



# Diversité, distribution spatiale et dynamique temporelle des petits eucaryotes dans des écosystèmes d'eau douce peu profond

Marianne Simon

## ► To cite this version:

Marianne Simon. Diversité, distribution spatiale et dynamique temporelle des petits eucaryotes dans des écosystèmes d'eau douce peu profond. Protistologie. Université Paris Sud - Paris XI, 2014. Français. NNT : 2014PA112200 . tel-01205723

**HAL Id: tel-01205723**

**<https://theses.hal.science/tel-01205723>**

Submitted on 27 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 426 :  
GÈNES GÉNOMES CELLULES

Laboratoire : *Ecologie Systématique Evolution (ESE)*

## THÈSE DE DOCTORAT

SCIENCES DE LA VIE ET DE LA SANTÉ

par

**Marianne Simon**

# **Diversité, distribution spatiale et dynamique temporelle des petits eucaryotes dans des écosystèmes d'eau douce peu profonds**

Date de soutenance : 26/09/2014

**Composition du jury :**

Directrice de thèse :  
Directeur de thèse :

Purificación LÓPEZ-GARCÍA  
Ludwig JARDILLIER

DR2 CNRS (Université Paris-Sud)  
MC (Université Paris-Sud)

Rapporteurs :

Jean-François HUMBERT  
Isabelle DOMAIZON

DR INRA (ENS Paris)  
DR INRA (INRA Thonon-les-bains)

Examineurs :

Ramon MASSANA  
Laure GUILLOU  
Didier CASANE

Chercheur (Institut de ciències del mar, Barcelone)  
DR2 CNRS (Station Biologique de Roscoff)  
Professeur (Laboratoire LEGS, CNRS)





# Table des Matières

---

## Introduction

---

<b>I. DIVERSITE DES EUCARYOTES UNICELLULAIRES.....</b>	<b>3</b>
<b>A- PETITE HISTOIRE DE L'ETUDE DES EUCARYOTES UNICELLULAIRES, DE LEUR DECOUVERTE AUX CLASSIFICATIONS ACTUELLES. ....</b>	<b>3</b>
1) LES PREMIERES OBSERVATIONS. ....	3
2) DU XVIII AU XXEME SIECLE, DE PLUS EN PLUS D'ORGANISMES DECRITS. ....	8
3) PETITS ORGANISMES ET GRANDS DESORDRES : INTEGRATION DES MICRO-ORGANISMES AUX CLASSIFICATIONS DU VIVANT. ....	9
4) L'APPORT DES DONNEES MOLECULAIRES. ....	13
<b>B- LA DIVERSITE DES EUCARYOTES UNICELLULAIRES A L'HEURE ACTUELLE. ....</b>	<b>16</b>
1) DECRIRE LA DIVERSITE AUJOURD'HUI. ....	16
2) LA DIVERSITE PHYLOGENETIQUE EUCARYOTE TELLE QU'ON LA CONNAIT ACTUELLEMENT. ....	21
3) DIVERSITE METABOLIQUE ET ECOLOGIQUE DES EUCARYOTES UNICELLULAIRES. ....	30
<b>C- LES PETITS EUCARYOTES. ....</b>	<b>32</b>
1) DIVERSITE. ....	32
2) ABONDANCE ET IMPORTANCE ECOLOGIQUE.....	34
<b>II. LES EUCARYOTES UNICELLULAIRES EN EAU DOUCE.....</b>	<b>35</b>
<b>A- LES MILIEUX D'EAU DOUCE.....</b>	<b>35</b>
1) DIFFERENTS TYPES D'ECOSYSTEMES D'EAU DOUCE. ....	35
2) LES MARES ET RUISSEAUX, DES ECOSYSTEMES PARTICULIERS ....	36
3) LES ECOSYSTEMES AQUATIQUES DANS LE PARC NATUREL REGIONAL DE LA HAUTE VALLEE DE CHEVREUSE .....	39
<b>B- DIVERSITE ET ASPECTS DE L'ECOLOGIE DES EUCARYOTES UNICELLULAIRES EN EAU DOUCE.....</b>	<b>41</b>
1) DIVERSITE DES MICRO-ORGANISMES EUCARYOTES EN EAU DOUCE. ....	41
2) REPARTITION GEOGRAPHIQUE DES EUCARYOTES UNICELLULAIRES EN EAU DOUCE. ....	43
3) DYNAMIQUE DES EUCARYOTES UNICELLULAIRES EN EAU DOUCE. ....	46
<b>C- LES MARES, DES ECOSYSTEMES PEU ETUDIES. ....</b>	<b>51</b>
<b>III. OBJECTIFS.....</b>	<b>53</b>
AMELIORER NOTRE CONNAISSANCE DE LA DIVERSITE DES PETITS EUCARYOTES.....	53
DECRIRE LA DYNAMIQUE TEMPORELLE DES COMMUNAUTES DE MICRO-ORGANISMES EUCARYOTES .....	54
IDENTIFIER DES FACTEURS POTENTIELLEMENT IMPLIQUES DANS LA DYNAMIQUE SPATIO-TEMPORELLE DES COMMUNAUTES D'EUCARYOTES MICROBIENS DANS LES MARES ET RUS. ....	54

## Matériel et méthodes

---

<b>I. ECHANTILLONNAGE.....</b>	<b>57</b>
<b>II. PREPARATION DES ECHANTILLONS .....</b>	<b>60</b>
<b>III. EXTRACTION DE L'ADN TOTAL .....</b>	<b>61</b>
<b>IV. AMPLIFICATION ET SEQUENÇAGE DES ADN 18S.....</b>	<b>61</b>
1) CLONAGE ET SEQUENÇAGE SANGER (DIVERSITE DES HAPTOPHYTES EN EAU DOUCE) .....	61
2) PYROSEQUENÇAGE 454 (SUIVI DE LA DIVERSITE GENERALE DES EUCARYOTES).....	62
<b>V. NETTOYAGE DES SEQUENCES, CONSTRUCTION DES OTU ET AFFILIATION .....</b>	<b>63</b>
1) TRAITEMENT DES SEQUENCES ISSUES DU SEQUENÇAGE SANGER .....	63
2) TRAITEMENT DES SEQUENCES ISSUES DU PYROSEQUENÇAGE 454 FLX TITANIUM .....	64
<b>VI. MESURES DES PARAMETRES PHYSICO-CHIMIQUES.....</b>	<b>67</b>
<b>VII. ANALYSE DES DONNEES ISSUES DU PYROSEQUENÇAGE 454 .....</b>	<b>68</b>
1) COMPARAISON DES COMMUNAUTES : ESTIMATION DU DEGRE DE RESSEMBLANCE ENTRE ECHANTILLONS.....	68
2) DETERMINATION DU NOMBRE D'OTU PARTAGEES PAR DIFFERENTS ECHANTILLONS .....	71
3) INDICES DE RICHESSE ET DE DIVERSITE.....	71
4) DETERMINATION DE L'ABONDANCE DES OTU .....	71
5) DETERMINATION DE LA FREQUENCE DES GROUPES FONCTIONNELS.....	72
6) COMPARAISON DES CARACTERISTIQUES PHYSICO-CHIMIQUES .....	72
7) IDENTIFICATION DE PARAMETRES EVENTUELLEMENT IMPLIQUES DANS LA DYNAMIQUE SPATIO-TEMPORELLE DES COMMUNAUTES ET DES ORGANISMES QUI LES COMPOSENT .....	73
<b>VIII. RECONSTRUCTIONS PHYLOGENETIQUES.....</b>	<b>74</b>

## Chapitre 1: Diversité des haptophytes en eau douce

---

<b>I. INTRODUCTION : RESUME DU CONTEXTE ET OBJECTIFS.....</b>	<b>77</b>
<b>II. MANUSCRIT DE L'ARTICLE 1 : " NEW HAPTOPHYTE LINEAGES AND MULTIPLE INDEPENDENT COLONISATIONS OF FRESHWATER ECOSYSTEMS" .....</b>	<b>78</b>
A- SUMMARY .....	79
B- INTRODUCTION .....	80
C- RESULTS AND DISCUSSION .....	83
D- SUPPLEMENTARY INFORMATION.....	92

<b>III. EPILOGUE : EXPLORATION DE LA DIVERSITE DES HAPTOPHYTES DANS QUATRE MARES ET UN RU PAR PYROSEQUENÇAGE 454 .....</b>	<b>99</b>
--	-----------

## Chapitre 2 : Exploration de la diversité des petits eucaryotes dans cinq écosystèmes d'eau douce peu profonds aux caractéristiques contrastées

---

<b>I. INTRODUCTION : RESUME DU CONTEXTE ET OBJECTIFS .....</b>	<b>107</b>
<b>II. MANUSCRIT DE L'ARTICLE 2 : "COMPLEX COMMUNITIES OF SMALL PROTISTS AND UNEXPECTED OCCURRENCE OF TYPICAL MARINE LINEAGES IN SHALLOW FRESHWATER SYSTEMS" .....</b>	<b>108</b>
A- SUMMARY .....	109
B- INTRODUCTION .....	110
C- RESULTS .....	112
D- DISCUSSION .....	124
E- EXPERIMENTAL PROCEDURES .....	128
F- SUPPLEMENTARY INFORMATION .....	132
<b>III. EPILOGUE .....</b>	<b>136</b>
A- UNE SURPRENANTE DIVERSITE DES MAST EN EAU DOUCE .....	136
B- DES COMMUNAUTES TRES PROCHES DANS DES ECHANTILLONS CORRESPONDANT THEORIQUEMENT A DES FRACTIONS DE TAILLE CELLULAIRE DIFFERENTES. ....	138

## Chapitre 3 : Diversité, écologie et dynamique des communautés de petits eucaryotes au cours d'un suivi mensuel sur deux ans dans quatre mares et un ru.

---

<b>I. INTRODUCTION : CONTEXTE ET OBJECTIFS .....</b>	<b>145</b>
<b>II. MANUSCRIT DE L'ARTICLE 3 : HECTIC TEMPORAL DYNAMICS OF COMPLEX SMALL EUKARYOTE COMMUNITIES IN SHALLOW FRESHWATER SYSTEMS .....</b>	<b>146</b>
A- ABSTRACT .....	147
B- INTRODUCTION .....	148
C- MATERIAL AND METHODS .....	150
D- RESULTS .....	153
E- DISCUSSION .....	164
F- CONCLUSIONS .....	169
G- SUPPLEMENTARY INFORMATION .....	170

## Chapitre 4 : Résilience rapide des communautés planctoniques de petits eucaryotes après sécheresse complète de leur écosystème.

---

<b>I. INTRODUCTION : CONTEXTE ET OBJECTIFS.....</b>	<b>183</b>
<b>II. MANUSCRIT DE L'ARTICLE 4 : FAST RECOVERY OF PLANKTONIC EUKARYOTE COMMUNITIES AFTER DROUGHT EVENTS IN TWO SHALLOW FRESHWATER ECOSYSTEMS. ....</b>	<b>184</b>
<b>A- ABSTRACT.....</b>	<b>185</b>
<b>B- INTRODUCTION .....</b>	<b>186</b>
<b>C- MATERIALS AND METHODS .....</b>	<b>187</b>
<b>D- RESULTS AND DISCUSSION .....</b>	<b>189</b>
<b>E- CONCLUSIONS .....</b>	<b>200</b>
<b>F- SUPPLEMENTARY INFORMATION .....</b>	<b>201</b>

### Discussion générale

---

<b>I. LA RICHESSE INEXPLOREE DES MARES ET RUISSEAUX .....</b>	<b>207</b>
<b>II. LA BARRIERE ENTRE MILIEU MARIN ET EAU DOUCE, MOINS INFRANCHISSABLE QUE PREVU? .....</b>	<b>209</b>
<b>III. DYNAMIQUE SPATIO-TEMPORELLE DES COMMUNAUTES DE PETITS EUCARYOTES DANS LES MARES ET RU .....</b>	<b>213</b>
<b>IV. COMPLEMENTARITE DU CLONAGE / SEQUENÇAGE SANGER ET DU PYROSEQUENÇAGE 454 .....</b>	<b>217</b>

### Conclusions

---

<b>I. LES PRINCIPALES INFORMATIONS DE CE TRAVAIL .....</b>	<b>223</b>
<b>II. QUELQUES PERSPECTIVES.....</b>	<b>224</b>

### Bibliographie

---

# Mille mercis...

Durant ma thèse, j'ai travaillé sur les *petits* eucaryotes des *petits* écosystèmes d'eau douce et de *petite* profondeur. S'il y a bien une chose en revanche qui n'est pas petite dans l'histoire, c'est les remerciements que je dois à tant de gens.

Merci tout d'abord aux membres du jury, Isabelle Domaizon, Jean-François Humbert, Didier Casane, Laure Guillou et Ramon Massana, merci d'avoir accepté d'évaluer mon travail, même si un manuscrit mi-avout, ce n'est pas un cadeau...

Merci ensuite à l'ED GGC et à ses responsables toujours disponibles. Merci au comité de thèse qui a bien voulu me donner ses conseils avisés à mi-parcours : Didier Casane, mon tuteur de thèse, Jean-François Humbert, Gérard Lacroix, et mes chefs.

Merci infiniment à mes chefs, Puri et Ludwig, de m'avoir accueillie parmi vous, merci d'avoir toujours été présents pour les questions (à toute heure!), merci pour les petites réunions à l'issue desquelles les choses se désembrument et prennent un sens, merci pour les dernières semaines (mois?) un peu plus intenses. J'ai, entre autres choses, décidé de tenter l'aventure de la thèse en pensant qu'il pouvait ne pas être inutile pour une prof d'avoir quitté l'école ne serait-ce qu'un moment avant d'y revenir de l'autre côté de l'estrader. Grâce à ces quelques années auprès de vous, je crois que mon enseignement ne sera pas le même, merci pour tout ce que vous m'avez appris. Gracias también por las cenas de primavera (y las tortillas tan deliciosas...), los aperitivos en el jardín... Pues, ¡muchas gracias por los últimos cuatro y media años!

Merci mille fois également à Gwendal, Philippe et David, pour votre aide infiniment précieuse, pour le traitement des séquences de 454, les amorces, les longues après-midi de questions de stat, les sorties échantillonnage, j'en passe et des meilleures. Merci pour tous vos conseils. Il est clair que ce travail ne serait pas là sans vous. Merci encore.

Pao, ton nom devrait être sur la première de couverture plutôt que dans la page des remerciements. Pour les 243 535 001 dosages de chloro, pour les levers aux aurores avant le terrain, pour le nombre de fois où tu as risqué ta vie en affrontant la jungle de fougères ou les mers de boue sans fond, le froid glacial ou les tiques, pour les filtrations / PCR / purifs / pools jusqu'au bout de la nuit. Mais merci aussi pour les restos, les dîners chez toi, la piscine, les confidences, les sorties théâtre... Pour les douches, les nuits chez toi, les transports jusqu'au rer. Pour les « C'est pâwti » et les bruitages quand on rince le zoo. Merci d'être toujours là pour tout le monde ; n'oublie pas qu'on est là pour toi aussi. Tu n'es pas près de te débarrasser de moi !

Merci ensuite à mes 2 co-bureaux, Thierry et Yann. Merci pour les moments de détente, les échanges de bons procédés (=sucreries), pour ce que j'ai pu apprendre avec vous (il ne me reste plus qu'à trouver comment caser « introgression », « *Pennisetum glaucum* », « précoce/tardif » et « promoteur de FT » dans mes conversations), pour les coups de main toujours proposés avec empressement. Le quotidien dans la salle T n'a jamais manqué de piquant. Yann, toutes mes excuses pour ne pas savoir arrêter de blablater, malgré tous mes efforts. C'est un honneur et un plaisir d'avoir partagé cette aventure avec toi, notamment le lent naufrage dans la spirale de la folie que nous avons vécu ces derniers mois. C'est rassurant de ne pas être seule dans cette déchéance. Everybody's kungfu fighting ! Mais ils ont beau dire, moi je crois que certains des dégâts psychologiques sont réversibles. Peut-être.

Merci à tous les gens qui font de ce labo un endroit agréable à vivre, et tout particulièrement aux microbios et aux habitants passés et présents du bureau 208. Merci à tous pour plus de choses que je n'en puis citer. Marie, Estelle et Charlie, merci pour tous vos conseils de paillasse. Marie, merci pour les échanges PBLV, j'ai encore rêvé d'une soirée karaoké. Estelle, merci pour ton accueil chaleureux dès mes premiers pas au labo, pour les mots fléchés au lit, les visioconférences transatlantiques. Charles, merci pour les réponses à toutes mes questions 454, pour les imitations de chauve-souris. Et tu sais, je ne t'ai

jamais vraiment cru quand tu disais passer de l'éthanol sur mes boîtes de clonage après que je les ai mises à l'étuve... **Jonathan**, merci pour tes histoires rocambolesques si bien mises en valeur par tes talents de conteur, et pour tes conseils cinématographiques (il faudra que tu me parles encore de « Tree of life »...). **Céline**, merci pour les questions de phylogénie, pour ton soutien perpétuel, les accompagnements-papotages en voiture, les gâteaux de pain. **Julien**, merci pour ton enthousiasme scientifique inspirant, pour les « petite Marianne » et pour me donner l'impression d'être utile en me laissant te fournir en nourriture ! **Boris**, notre premier grand garçon et chanteur préféré, merci de ne pas nous humilier aux soirées karaoké, d'avoir toujours un mot gentil. **Charlotte**, notre vie commune ne date pas de la thèse, mais ça a été un grand plaisir de te retrouver au labo et de partager cette histoire-là aussi avec toi, merci aussi pour ton aide précieuse pour le monitorat et la mission doctorale. **Aurélien**, petit garçon, petit poulain, merci pour l'entrain de ta jeunesse, merci de nous rappeler sans cesse notre grand âge, merci de tes impertinences (quoi que...). Merci à tous les participants des pauses horoscopes et le saint du jour de 16h (merci aussi donc à D. Genolini, et tout particulièrement à **Ste Rolande**, patronne de la salle T). Merci à vous tous et à **Puri**, **David**, **Philippe**, **Pao**, **Ludwig**, **Gwendal**, **Emmanuelle** (merci pour ton aide souriante en toute circonstance !), **Aloïdie** (première co-bureau, ça ne s'oublie pas), merci à **Alain**, **Najat**, **Jacqui**, **Kader**, **Gisèle**, **Damien**, **Antoine**, **Roxane**, **Eloïdie**, **Amandine**, **JT**, **Odile** et **Odile**, **Véro**, **Françoise**, **Sophie**, **Christian**, **Béatrice**, et à tant d'autres, merci à tous pour les labmeetings enrichissants, les déjeuners en salle café, les ateliers de préparation des « ESE special gift », les ateliers tricots, les soirées / apéros, les barbecues, les discussions.

Merci à tous les enseignants aux côtés desquels j'ai pu faire mes premiers pas en tant que "prof" lors de mon monitorat. Merci à toutes les équipes enseignantes de la prépa agreg, de Meth191, Biol102 et L2 Bioconcours, et particulièrement à **Hélène Vincent**, **Martine Thomas**, **Line Duportet**, **Solange Bertrand** et **Morgane Locker**, responsables de ces enseignements. J'ai beaucoup appris à votre contact, c'est une chance d'avoir pu rentrer en douceur dans le métier en étant si bien entourée. Merci aussi à tous les étudiants qui ont essuyé les plâtres...

Merci à tous les gens que j'ai pu croiser au cours de ma mission doctorale de 4<sup>ème</sup> année, et spécialement à **Sophie** et **Valérie**. C'était une expérience agréable et enrichissante.

Merci à mes cahiers petit format petits carreaux à spirales. De tous les journaux intimes que j'ai débutés dans ma vie, aucun n'aura été rempli avec une telle assiduité pendant si longtemps. Merci d'avoir été une version améliorée et externalisée de ma mémoire pour ma 4<sup>ème</sup> année.

Merci à tous les participants des Red 11 et Red 12, et bien sûr tout particulièrement à **Muriel** et **Hervé**. Ces exo-rencontres resteront un de mes très beaux souvenirs de thèse, pour l'ouverture qu'elles m'auront apportée vers d'autres disciplines un peu loin de moi depuis quelques années, pour m'avoir permis de voir (et toucher !) des komatiites et leurs olivines spinifex (ça existe donc vraiment !), et pour toutes les jolies rencontres, oiselières, genettesques et humaines.

En parlant de jolies exo-rencontres, un merci tout particulier aux résidents du mobile-home idéal. **Axel**, **Christophe**, **Claire**, **Clément**, **David**, **Fanny** et **Romuald**, merci à tous et à chacun pour les pâtes au pesto vegan, les devinettes, les confidences, les pestacles, et les *quelques* mails échangés, traits d'union entre les WE-parenthèses : j'espère qu'ils continueront un moment à ponctuer ma vie !

Merci au **rer B**. On te maudit souvent, tes oreilles doivent régulièrement siffler (ah mais c'est donc ça ! J'ai toujours cru que c'était pour signaler que les portes se fermaient...). Bien sûr les petites feintes du type changements d'horaire ou de trajectoire inopinés sont assez irritantes, et je ne te parle pas des quelques soirées passées dans les transports pour cause de colis suspect... Mais tu es aussi mon rendez-vous quotidien depuis 8 ans, et finalement tu m'as bien moins laissée sur le carreau que tu ne m'as gentiment transportée. Et grâce à toi j'ai eu le temps de bouquiner !

Merci enfin à tous ceux qui ont gentiment accepté de relire un bout de ce manuscrit, mes chefs bien sûr, et **Céline**, **Papa**, **Maman**, **Aurélien** le petit garçon. Merci mon **François** pour l'impression by night...

Et puis il y a aussi tous ceux qui n'ont pas tant à voir avec ma thèse, mais on prend si peu l'occasion de remercier...

A Maman et Papa, mes plus belles réussites (©MyriamGibert, 1993). Merci infiniment d'être toujours là, chacun et tous deux. Au risque de tomber dans le registre cadeau de fête des mères / pères : vous êtes sans nul doute les meilleurs parents du monde. Merci pour les WE changements d'air ces dernières années quand j'ai pu rentrer à mes maisons, les petits coups de fils (même si je ne réponds jamais, oui je sais!), tout ça. Alors merci encore à ma Maman (parce qu'on n'a qu'une maman, et cette maman, c'est sa maman) et à mon Papa UEP (parce que par définition un papa UEP...). Merci aussi à Aude et Philippe, de si bien vous occuper d'eux! ☺

Merci à toute ma fratrie aux degrés de parentèle variés. Quentin, même si tu étais positivement insupportable, je suis bien contente que tu sois celui avec qui j'ai partagé l'intégralité des 1ères années de ma longue vie. BenJ, qui c'est le bébé d'amour à sa Yaya? Merci ensuite à Mamie et Yoyo, pour tous les moments de complicité. Merci aussi à mes tontons Lulu et Rilou, à Kouka ma tata et à tatAgnès, à mes petits cousins (avec une pensée spéciale pour Maud!), à mes grand-pères et toute la famille. Je vous vois rarement mais c'est bien de vous savoir là.

Merci ensuite à Francine et Pierre et à toute la famille Huré-Bobot. Merci pour votre accueil chaleureux parmi vous. Merci pour les noëls, les vacances, les WE, les conseils EducNat!

Merci enfin à tous les amis que j'ai eu le bonheur de rencontrer dans ma vie. Delphine, Laura au collège, Sophie au lycée, Blandine, Marina, Lauren, Camille, Carole, Julia en prépa, AnneSo (merci pour ton soutien de co-fin de thèse!) et Chou, Hélène (merci pour les piscines chaque semaine sans exception (j'écris ça comme ça peut être que dans quelques années en le relisant on y croira...)) et Yan, Tatiana et Guigui (merci infiniment pour toute ton aide en agreg, en monitorat... je ne te remercierai jamais assez), Céline et Picpic, Marinouchka, et tous les bios (parce que la bio c'est beau) pendant les années Cachan, Charlotte; Lionel et Clélie, Agathe et Lapin, Claire (merci mille milliard de fois pour tes conseils de stat!), Hector et Valentin, Henri, Claire (merci de m'avoir si bien chouchoutée pour ma prise de poste!), Pierret' et Céline en agreg, Gwénaelle et Alban en M2, Hélène et Juliette en thèse (et tous les autres bien sûr, mais vous avez déjà eu votre paragraphe, ne soyez pas trop gourmands). On m'a plusieurs fois demandé durant ma carrière de petite intello si je n'en avais pas marre de passer tant de temps dans mes cahiers, et si je ne regrettais pas parfois de ne pas avoir pris une voie un peu plus cool. Chaque choix que j'ai fait dans mes études m'a permis de rencontrer certains d'entre vous, et ne serait-ce que pour cette raison, je ne regrette aucune nuit trop courte, aucun WE enfermée!

Enfin, last but not least, merci beaucoup infiniment à François. Au moment où j'écris ces lignes tu es patiemment assis à côté de moi au labo en attendant que je sois prête à imprimer... Pauvre de toi! Merci infiniment pour ta patience, pour ton soutien logistique et moral ces dernières semaines (si ce n'est plus...). Promis, je vais tâcher de perdre au plus vite l'habitude rapidement prise que tu t'occupes de moi comme d'une princesse en chaque instant (je ne suis pas bien sûre d'avoir une résilience très élevée dans ce domaine, mais je ferai des efforts!). Merci d'avoir été à mes côtés dans ce moment-là comme tu l'es toujours quand j'en ai besoin. <3 Un merci grand comme jusqu'à la lune et retour pour ces bientôt 8 années de vie commune, et surtout pour toutes celles qui suivront. Je les imagine, ma foi, fort agréables...

Je crois que je vais m'arrêter là, malgré tout ce et ceux dont je n'ai pas parlé, c'est déjà bien trop long, comme à mon habitude. Pourtant, comme dirait mon idôle (non, pas les mitochondries, mon autre lubie) "*J'ai l'âme lourde encor d'amour inexprimée*"... ☺





# Introduction

---



*"Se donner du mal pour les petites choses, c'est parvenir aux grandes, avec le temps."*  
Samuel Beckett

**Illustration :** Petit échantillon de la diversité de formes et de taille au sein des eucaryotes unicellulaires, représentés à l'échelle à côté d'une tête d'épingle. (Finlay, 2002)



# Introduction

---

*"In the clearest waters, and also in the troubled, strongly acid, and salt fluids of the various zones of the earth ; in springs, rivers, lakes, and seas ; in the internal moisture of living plants and animal bodies, and, at times, carried about in the vapor and dust of the whole atmosphere of the earth, exists a world, by the common senses of mankind unperceived, of very minute living beings."*

Pritchard, *History of Infusory*, 1852

## I. Diversité des eucaryotes unicellulaires.

*"L'essentiel est invisible pour les yeux"*

Antoine de St Exupéry

### A- Petite histoire de l'étude des eucaryotes unicellulaires, de leur découverte aux classifications actuelles.

#### 1) Les premières observations.

##### *Préambule : particulae illoae insensibilis.*

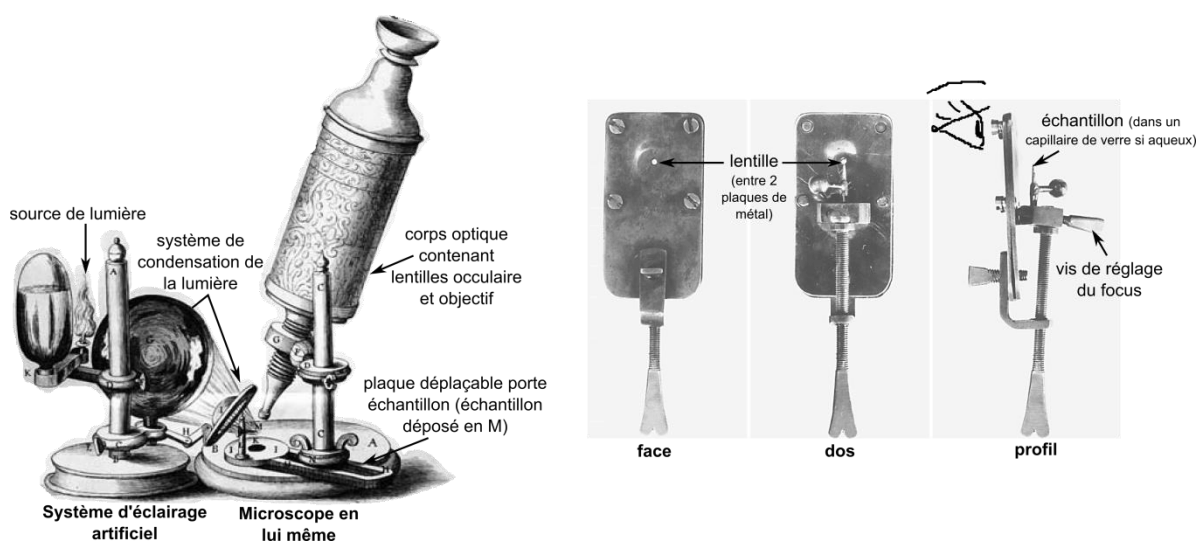
Au XV<sup>ème</sup> siècle, avant l'utilisation des microscopes, on ne connaissait du vivant que les organismes visibles à l'œil nu – certains animaux, végétaux et champignons. Pourtant, la présence des microorganismes dans le quotidien des humains a un impact bien visible. Fracastorius (v. 1480–1553), un savant italien, s'est particulièrement intéressé aux maladies contagieuses et notamment la syphilis, sur laquelle il a composé de nombreux « poèmes médicaux » pour lesquels il est célèbre aujourd'hui encore. On lui doit notamment la première description scientifique de la nature de la contagion et de l'infection, et des théories quant à l'origine de ces phénomènes (Fracastoro 1546). Il suppose notamment l'existence de germes qu'il nomme *seminaria contagionum*, et qu'il définit comme des "particules trop petites pour être appréhendées par nos sens" ("*particulae illoae insensibilis*") mais qui, dans un milieu approprié, sont capables de se reproduire et donc d'infecter les tissus alentours » (Garrison 1910).

Pendant plus d'un siècle, l'existence de tels organismes microscopiques reste hypothétique. Jusqu'à ce que l'utilisation scientifique de microscopes permette leur observation.

**La première description d'un microorganisme publiée : Robert Hooke 1665.**

Robert Hooke (1635 - 1703) est un grand scientifique anglais de la fin du XVII<sup>ème</sup> siècle. Le début de sa vie est consacré à des études d'art. Après avoir été apprenti chez un grand portraitiste, il entre comme *undergraduate* à l'Oxford's Christ Church College dont il sort vraisemblablement sans diplôme. Ces années à Oxford lui auront pourtant été profitables puisqu'il y rencontre les futurs fondateurs de la Royal Society de Londres, société à laquelle il participe dès sa création en 1660. En 1662, il est nommé *Curator of Experiments* de cette Royal Society, charge qu'il exerce jusqu'en 1677. A ce titre, il est chargé de toutes les expériences réalisées durant les réunions hebdomadaires de la société. Grâce à ses nombreux travaux dans tous les domaines de la science, Hooke est nommé secrétaire de la Royal Society en 1677 (Barcat 2003; Gest 2004a; b).

Alors qu'il est *Curator of Experiments*, il réalise de nombreuses observations microscopiques à l'aide d'un microscope assez complexe, à deux lentilles et en lumière réfléchie, avec un système d'éclairage artificiel de sa conception (Figure 1).



**Figure 1 : Les microscopes utilisés par Hooke (gauche) et Leeuwenhoek (droite).** Hooke décrit précisément dans *Micrographia* les petits microscopes à une lentille et lumière transmise comme ceux de Leeuwenhoek, ainsi que leur conception, avant que Leeuwenhoek ne les utilise. Il reconnaît ne pas les utiliser malgré leur meilleure résolution par ce qu'ils sont mauvais pour ses yeux (Gest 2004b). Adapté de Corliss, 2002; Donaldson, 2010.

Hooke présente les résultats de ses observations lors de réunions de la Royal Society en avril 1663. Parmi ses nombreux objets d'étude se trouvent des tissus végétaux, dont du liège – à partir duquel il est le premier à proposer le terme de « cellules », et une moisissure. Il décrit

également ses observations dans *Micrographia*, édité pour la première fois en 1665 (Hooke 1665). Cet ouvrage, le premier dédié à des observations microscopiques, contient 60 observations d'objets variés, dont de nombreux objets biologiques. Toutes sont accompagnées d'illustrations fines et précises que l'on doit certainement au passé d'artiste peintre de l'auteur. On y trouve notamment la description d'une « *Blue Mould* », moisissure trouvée sur la couverture en cuir de mouton rouge d'un petit livre (Hooke 1665; Gest 2004b). Cette description très précise, accompagnée d'un dessin d'observation, a permis la détermination ultérieure de l'organisme en question : il s'agit d'un champignon du genre *Mucor* (Figure 2). La première description certaine d'un microorganisme est donc celle d'un eucaryote microscopique !



**Figure 2: La première description publiée d'un micro-organisme: le champignon *Mucor*.** A gauche, le dessin d'observation de Hooke (Observation XX, *Micrographia*, 1665). On trouve la description de cette figure dans le texte correspondant : "multitudes of small long cylindrical and transparent stalks, not exactly streight, but a little bended with the weight of a round and white knob that grew on the top of each of them; many of these knobs I observ'd to be very round and of a smooth surface, such as A A, etc. others smooth likewise, but a little oblong, as B.; several of them a little broken, or cloven with chops at the top, as C; others flitter'd as 'twere, or flown all to pieces, as DD." A droite, *Mucor plumbens* observé en microscopie optique (photographie de la Truman State University's microfungi image gallery, <http://microfungi.truman.edu>). Les lettres ont été ajoutées sur la photographie en faisant le parallèle avec le dessin d'observation et la description de Hooke.

### **Les travaux de van Leeuwenhoek.**

Antony van Leeuwenhoek (Delft, 1632 – 1723) est un scientifique au parcours peu commun. Alors que de nombreux scientifiques de l'époque étaient issus de familles aisées qui leur permettaient de réaliser des études et recherches sans se soucier de gagner leur pain, Leeuwenhoek quitte l'école vers 15 – 16 ans pour apprendre le métier de drapier. Il ouvre sa

## Introduction

propre boutique à Delft (Pays-Bas) en 1654. Sans doute a-t-il commencé à utiliser le pouvoir grossissant des lentilles dans son métier de drapier, pour vérifier la qualité des tissus (Gest 2004b). Il se passionne pour les observations microscopiques, et réalise bientôt des lentilles grossissantes de très bonne qualité. Il utilise alors des microscopes simples à une lentille et à lumière transmise (Figure 1) pour réaliser des observations variées. Il partage ses observations avec des amis, dont Reinier de Graaf, célèbre anatomiste résidant également à Delft. En 1673, de Graaf écrit au secrétaire de la Royal Society de Londres pour attirer son attention sur les travaux de Leeuwenhoek. Dans l'année, une délégation de la Royal Society visite Delft et rentre à Londres très impressionnée par l'efficacité des petits microscopes de Leeuwenhoek et par l'utilisation faite par leur concepteur (Corliss 2002). Suite à cette visite, Leeuwenhoek enverra toute sa vie la description de ses observations au secrétaire de la Royal Society (qui, rappelons-le, n'est autre que Hooke à partir de 1677). Cette correspondance étalée entre 1673-1674 et 1723 sera composée de quelques 200 longues lettres, en néerlandais, sur des sujets très variés (Corliss 2002).

On doit en grande partie notre connaissance actuelle des travaux de Leeuwenhoek à Clifford Dobell, protistologue anglais du début du XX<sup>ème</sup> siècle et fervent admirateur du « drapier de Delft », qui a écrit sa biographie et traduit, parfois pour la première fois, une grande partie de ses lettres à la Royal Society (Dobell 1932). Dobell a notamment mis en évidence une première mention à des observations de micro-organismes eucaryotes dès l'une des toutes premières lettres à la Royal Society. En effet, la lettre du 7 septembre 1674, évoque l'observation d'organismes microscopiques, que Leeuwenhoek appelle « animalcules », dans des mares d'eau douce près de Delft. Le texte consacré à cette observation est assez court. Les descriptions morphologiques sont encore peu précises, Leeuwenhoek parle surtout de la diversité observée, de la petite taille des organismes et de leurs mouvements, caractéristiques qui l'impressionnent particulièrement. D'après Dobell, Leeuwenhoek aurait alors observé des spirogyres, des rotifères et des ciliés non identifiables (Dobell 1932).

Deux ans plus tard, Leeuwenhoek fait de nouveau référence à des observations de micro-organismes. La longue lettre n°18 datée du 9 octobre 1676 est en effet entièrement consacrée à ses observations « d'animalcules ». Pour cette raison et sans doute parce que des extraits de cette lettre ont été traduits et publiés par la Royal Society peu de temps après son arrivée à Londres, cette lettre a longtemps été considérée comme la première mention de micro-organismes par Leeuwenhoek ; elle est aussi connue sous le nom de "*Letter on the protozoa*". Pour convaincre les sceptiques, Leeuwenhoek retranscrit dans cette lettre de longs



passages de son « journal », l'équivalent de nos cahiers de manipulations actuels. Il en résulte que cette lettre contient des descriptions détaillées de ses observations au jour le jour, mais aussi de ses protocoles. Il y décrit 6 observations dans de l'eau de pluie collectée et conservée de façons diverses, une observation d'eau de rivière, une de l'eau de son puits, une en eau de mer, une dans du vinaigre, et de nombreuses observations d'infusions de différents composés (poivre en grain ou moulu – la célèbre « *pepper water* », gingembre, vêtements...).

Il me semble que cette lettre n°18 contient quatre grands types de données principales :

- Cette lettre constitue la première description non discutée de bactéries, qu'il qualifie de « *very little animalcules* ».
- Leeuwenhoek y décrit de nombreux microorganismes eucaryotes, dont beaucoup de ciliés : il parvient même à en voir les cils<sup>1</sup>. Sans aucune illustration, certaines descriptions sont assez précises pour permettre une reconnaissance du genre voire de l'espèce observée. Dobell recense ainsi les descriptions de *Vorticella* (cilié), *Spumella* (que Dobell nomme *Monas*, chrysophyte), *Bodo caudatus* (Kinétoplastidés, Euglenozoa)... (Dobell 1932)
- Leeuwenhoek est l'un des premiers à donner des indications de taille précises. Ce n'est pas une mince affaire, puisqu'il n'a à sa disposition ni outil ni même unité de mesure adaptés. En effet, la plus petite unité de mesure à Delft en cette fin de XVII<sup>ème</sup> siècle est le inch, qui à l'époque équivaut à 3,29 cm. Il mesure donc de petits objets mesurables sans grossissement (grains de sable, œil de mouche, cheveux...) grâce à une « règle » sur laquelle il a divisé chaque inch en dixièmes. Il indique ensuite la taille des animalcules en fraction de la taille de ces petits objets, et ses estimations sont souvent de bonnes approximations de la taille réelle (Corliss 2002). La lettre 18 regorge de ces comparaisons très imagées telles que “they were no bigger, through my microscope, than very little sand-grains to one's naked eye; and I must say, that I deem them to be a good three or four times smaller than the thickness of the hair of a mite, or one of the little hairs wherewith the body of a louse is beset.” (Description d'animalcules observés dans une infusion de noix de muscade).
- Le dernier point à relever dans cette lettre 18, c'est que Leeuwenhoek ne se contente pas de décrire la morphologie des organismes qu'il observe. Il note également des observations sur leur comportement (e.g. type de nage, coiling / uncoiling du pied de *Vorticella*) et leur écologie : il observe une dynamique saisonnière (dans son puits, les animalcules ne sont visibles qu'en été), le remplacement de communautés par d'autres, la sensibilité aux particularités du milieu.

---

<sup>1</sup> “thin little feet, or little legs, which were moved very nimbly, and which I was able to discover only after sundry great efforts, and wherewith they brought off incredibly quick motions.”



## Introduction

Ces observations sont accueillies avec beaucoup de scepticisme, notamment à la Royal Society. Hooke répète lui-même les observations de Leeuwenhoek sur des infusions diverses. Le 15 novembre 1677, il présente une infusion de grains de poivre entiers dans de l'eau de pluie, infusée plusieurs jours au cours desquels il a pu quotidiennement observer des animalcules. Ces observations sont confirmées par différentes personnes le jour de la réunion, assurant ainsi la crédibilité des découvertes de Leeuwenhoek auprès de la Royal Society, qui l'élit parmi ses membres en 1680 (Gest 2004a).

Leeuwenhoek continue ses observations microscopiques toute sa vie (Dobell 1932). Il sera le premier à observer des microorganismes parasites. Au tournant du XVII<sup>ème</sup> au XVIII<sup>ème</sup> siècle, alors qu'il a presque 70 ans, il fait encore de grandes découvertes, surtout d'eucaryotes aquatiques libres : *Volvox* et un test (coquille) de foraminifère début 1700, *Chlamydomonas* en 1702... Outre les descriptions morphologiques, il décrit par exemple la conjugaison chez des ciliés (9 novembre 1695), la reproduction de *Volvox* (janvier 1700).

A la même époque, Louis Joblot (1645-1723), professeur de mathématiques à l'Académie Royale de Peinture et de Sculpture de Paris, consacre une grande partie de son temps à développer de nouveaux microscopes. Grâce à ces outils, il observe de nombreux insectes mais également des micro-organismes présents dans du vinaigre ou diverses infusions. Ses travaux sont publiés en 1717 dans l'ouvrage *Defcriptions de plufieurs nouveaux Microfcopes, avec des Obfervations fur une grande multitude d'Infectes qui naiffent dans les liqueurs, &c.* puis de façon plus approfondie dans *Obfervations d'hiftoire naturelle faites avec le Microfcope, Sur un grand nombre d'Infectes, & fur les Animalcules qui fe trouvent dans les liqueurs préparées & dans celles qui ne le font pas*. La seconde partie de dernier ouvrage notamment, publié en 1754 de façon posthume, contient de nombreuses descriptions et planches dessinées d'animaux microscopiques (les anguilles du vinaigre surtout) et de micro-organismes.

## 2) Du XVIII au XX<sup>ème</sup> siècle, de plus en plus d'organismes décrits.

Après ces premières observations, de nouvelles espèces de micro-organismes sont décrites tout au long du XVIII<sup>ème</sup> siècle, parmi les ciliés (*Paramecium*, Joblot 1718), les dinoflagellés (*Noctiluca*, Baker 1753) et d'autres taxa. Otto Friederich Müller est considéré comme le premier à avoir proposé une taxonomie organisée des micro-organismes. Dans une publication posthume (1786), il nomme, décrit et présente des dessins de près de 300 espèces

de micro-organismes eucaryotes. Certains des noms de genres qu'il propose sont encore utilisés aujourd'hui (e.g. *Paramecium*, *Colpoda*, *Cyclidium*) (Corliss 1991).

L'étude des micro-organismes prend plus d'ampleur au cours du XIX<sup>ème</sup> siècle alors que les microscopes, toujours plus performants, prennent une place croissante dans la vie des biologistes. Avec l'augmentation du nombre d'espèces décrites vient une volonté croissante de classer les organismes dans des groupes de différents rangs taxonomiques, comme en témoignent plusieurs ouvrages de systématique des « protozoaires » (e.g. W.S. Kent 1880-1882 ou O. Bütschli 1880-1889 (Corliss 1991; Fokin 2013). Au cours de ce siècle, de plus en plus de scientifiques, dans de plus en plus de régions du monde, se penchent sur la diversité et la biologie des micro-organismes. Ils commencent à se spécialiser dans l'étude de groupes précis, les ciliés semblant souvent concentrer l'intérêt (Cf. travaux de Balbiani, Chatton, Fauré-Fremiet (J. Corliss, 1991)).

Ces travaux mettent en évidence une diversité de morphologies et modes de vie, qui ne cadrent pas avec l'image du vivant prépondérante depuis des siècles, et vont pousser les naturalistes de l'époque à bouleverser un cadre établi depuis bien longtemps.

### **3) Petits organismes et grands désordres : intégration des micro-organismes aux classifications du vivant.**

#### ***Les prémices des classifications du vivant.***

Depuis l'Antiquité, les savants cherchent à organiser le vivant. On pense qu'Aristote (384-322 av. J.C.) a, le premier, essayé de classer les organismes (Duris & Gohau 2011). S'il s'est presque exclusivement concentré sur les animaux, son disciple Théophraste (v.376-v.288 av. JC) s'est quant à lui attelé à la classification des végétaux. Ce dernier est de plus à l'origine de la première distinction explicite entre le monde animal et le monde végétal. À partir de là, la nature est découpée en trois règnes : animal, végétal et minéral. Dans les siècles qui suivent, les savants occidentaux se désintéressent quelque peu des sciences naturelles et, à quelques exceptions près, on ne trouve plus de description ou tentative de classification très approfondie de la diversité du vivant. La Renaissance en revanche marque le retour de ces études, et différentes classifications du monde vivant sont proposées, sur la base de critères très variés comme la morphologie, l'écologie, l'utilité pour l'homme ou l'ordre alphabétique, mais qui toutes conservent la dichotomie animal / végétal. D'après Duris et Gohau, les toutes premières classifications sont en général volontairement artificielles. Elles auraient surtout un rôle pratique, pédagogique, plus qu'une volonté de révéler un « ordre » dans la nature qui pourrait préexister aux classifications humaines.





## Introduction

Plus tard, ce point de vue sur les classifications du vivant change, notamment avec Linné (1707-1778). Pour lui, décrire, classer et nommer les « corps naturels » est une tâche de première importance, le pilier de toutes les sciences : une « science divine » (Duris & Gohau 2011). Il ne s'agit plus de trouver un cadre pédagogique qui soit pratique pour discuter, mais de révéler un ordre préexistant de la nature, créé par dieu et non inventé par l'homme. Ainsi, dans le *Systema naturae* (1735 pour la première édition), il propose une classification détaillée des trois règnes de la nature : les minéraux, les végétaux et les animaux. Lui-même ne s'attelle pas à la classification des micro-organismes – les animalcules – qu'il regroupe pour la plupart dans un genre *Chaos infusoria* (Giordan 1987), ignorant différents genres déjà définis à l'époque (e.g. *Cyclidium*, *Paramecium* (Corliss 1991)). Mais son ouvrage offre une méthodologie qui sera utilisée par la suite pour les organismes unicellulaires comme pour les macro-organismes. Cette méthodologie consiste en la détermination d'un certain nombre de caractères morphologiques précis, explicitement définis, et de classer les organismes selon la « version » de ces caractères qu'ils possèdent.

### ***Ni animaux ni végétaux : création d'un 3<sup>ème</sup> règne du vivant.***

Dès la seconde moitié du XVIII<sup>ème</sup> siècle, les scientifiques entreprennent d'intégrer au cadre existant les micro-organismes décrits, en se fondant sur différentes caractéristiques morphologiques et physiologiques. Les « animalcules » photosynthétiques immobiles sont alors regroupés avec les plantes terrestres dans le règne végétal tandis que celles qui sont hétérotrophes et mobiles rejoignent le règne animal. Mais les choses ne sont pas si simples... De nombreux micro-organismes associent les critères de mobilité et de type trophique dans des combinaisons qui ne correspondent ni à celle des animaux ni à celle des végétaux, de sorte qu'ils ne peuvent être aisément classés dans aucun des règnes existant (Whittaker 1969). Certains, comme *Euglena*, se retrouvent alors réclamés à la fois par les botanistes et les zoologistes.

Face à ces incompatibilités, plusieurs scientifiques proposent au cours du XIX<sup>ème</sup> siècle la création d'un troisième règne de la nature (ou quatrième lorsqu'ils prennent en compte le règne minéral). Ce nouveau règne change de nom, de définition et de composition exacte selon les auteurs considérés. En règle générale, ses contours restent très flous, et il est souvent défini par exclusion, pour y classer les organismes qui ne répondent pas aux définitions que l'on donne des autres règnes. Pour certains, il regroupe presque exclusivement des organismes unicellulaires, pour de nombreux autres, il sert de « fourre-tout » où l'on retrouve la plupart des unicellulaires mais aussi des macro-organismes. Les « éponges », les « algues », les champignons et les bactéries semblent être les groupes les plus

problématiques ; ils se déplacent régulièrement d'un règne à l'autre dans les différentes classifications. De toutes les propositions énoncées dans la seconde moitié du XIX<sup>ème</sup> siècle, c'est sans nulle doute le règne Protista<sup>2</sup> d'Haeckel (1866 pour sa première version) qui est le plus resté dans les mémoires, mais plusieurs l'ont précédé qui ont aussi participé à l'évolution de notre vision du vivant : les règnes Protozoa d'Owen (1858), Primigenum de Hogg (1860), Primalia de Cassin et Wilson (1863) [Tableau 1 (Corliss 1998; Scamardella 1999)] par exemple. Il faut néanmoins reconnaître à Haeckel plusieurs innovations. Tout d'abord il abandonne les minéraux en tant que règne. Ensuite, depuis la publication par Darwin de *l'Origine des espèces* en 1859, les classifications ont acquis une nouvelle perspective, un aspect historique. De plus en plus on veut créer des groupes qui aient un sens du point de vue des relations de parenté entre organismes. Haeckel soutient fermement cette idée et semble être le premier à proposer un arbre du vivant, montrant les relations entre les différents organismes (Tableau 1). Il reconnaît d'autre part le côté artificiel du regroupement des organismes unicellulaires en un seul règne. L'arbre qu'il propose montre de plus qu'il considère son règne Protista comme l'équivalent des règnes Plantae et Animalia, alors que pour d'autres scientifiques contemporains comme Hogg (1860) ce nouveau règne était plutôt une zone de continuité entre les animaux et les végétaux. Finalement, dans ce premier arbre (1866) les Protista paraissent dériver d'un ancêtre partagé avec les deux autres règnes, et non être eux-mêmes les ancêtres des macro-organismes. Cette dernière idée ne sera pas acceptée avant longtemps, et Haeckel lui-même présente en 1879 un arbre de l'ascendance de l'homme montrant les bactéries, les amibes et d'autres micro-organismes hétérotrophes comme les plus vieux ancêtres de l'homme.

### ***La multiplication des règnes.***

Après l'abandon du dogme de la dichotomie Animal/Végétal, un nouveau grand pas est fait dans l'évolution des classifications du vivant, avec la mise en évidence d'une différence d'organisation cellulaire fondamentale entre deux groupes d'êtres vivants. En effet, Edouard Chatton (1883–1947) note en 1937 qu'il existe deux grands types de cellules : les cellules avec un noyau et d'autres compartiments cellulaires délimités par une membrane (réticulums, mitochondries...), qu'il nomme *eucaryotes*, et des cellules sans noyau ou *procaryotes*. Alors que cette distinction eucaryotes / procaryotes ne deviendra populaire que quelques vingt années plus tard, Copeland propose dès 1938 de l'introduire dans la

<sup>2</sup> Aujourd'hui, le terme « protistes » dérivé du règne proposé par Haeckel dans la seconde moitié du XIX<sup>ème</sup> siècle est encore couramment utilisé, non dans son sens originel mais pour désigner l'ensemble des eucaryotes unicellulaires (parfois à l'exclusion des champignons). Compte-tenu du sens historique précis qu'il porte, et du peu de gain de praticité qu'il apporte par rapport à un énoncé plus exact comme « micro-organismes eucaryotes », il me semble discutable d'en encore utiliser ce terme.

## Introduction

classification du vivant. Il sort alors les bactéries (dont les cyanobactéries) des règnes existant, et les regroupe dans un quatrième règne à part entière : Monera, qui chez lui est clairement défini, et contient toutes les cellules anucléées, appelées alors monères, et seulement elles (Tableau 1). Le vivant est alors découpé en quatre grands règnes.

Quelques années plus tard, Whittaker observe parmi les organismes pluricellulaires trois grands modes de nutrition : par ingestion chez les animaux, par photosynthèse chez les végétaux ou par absorption chez les champignons. Ces trois modes de nutrition impliquant de nombreuses adaptations morphologiques et physiologiques, elles doivent selon lui être associées à trois histoires évolutives différentes. Il propose donc la création d'un nouveau règne, le règne Fungi, défini par une nutrition par absorption, et dans lequel il regroupe les actuels champignons et oomycètes (1957). Les règnes Animalia et Plantae sont alors définis par leur nutrition par ingestion et photosynthèse respectivement. Définir ainsi les règnes Fungi, Animalia et Plantae lui permet également de ne conserver que les organismes unicellulaires dans son règne Protista, et donc d'en proposer une définition plus claire que celles proposées pour les règnes mixtes. En 1969, il adopte le règne Monera de Copeland, et présente donc une classification du vivant en cinq règnes (Tableau 1). Whittaker reconnaît que ses trois règnes pluricellulaires restent polyphylétiques (Whittaker 1969). Cependant, créer uniquement des groupes monophylétiques nécessiterait une forte augmentation du nombre de règnes et donc de la complexité de la classification, or « *Monophyly is a principal value of systematics, but like other values is not absolute and will not always be followed to the sacrifice of other objectives* » (Whittaker 1969).

Ainsi, au milieu du XX<sup>ème</sup> siècle, il semble admis que les eucaryotes unicellulaires ne sont ni des plantes, ni des animaux, et sont également très différents des bactéries. A de rares exceptions près, ils sont toujours considérés comme les ancêtres des eucaryotes pluricellulaires. Enfin, malgré une volonté croissante de créer des groupes en fonction des liens de parenté entre organismes, tous les règnes des classifications proposées à l'époque sont polyphylétiques. Notamment, tous les eucaryotes unicellulaires sont groupés dans un même règne, alors que l'incapacité des scientifiques à trouver un caractère partagé qui permettrait de définir ce règne les a déjà conduits à penser qu'il regroupait des organismes très différents. Cependant, les techniques disponibles en ce milieu de XX<sup>ème</sup> siècle ne permettent pas d'accéder à suffisamment de caractères discriminants pour déterminer les relations de parentés entre eucaryotes unicellulaires. En effet, toutes les classifications sont alors basées sur des caractéristiques écologiques, physiologiques ou morphologiques observées par microscopie optique. Or, les caractéristiques écologiques résultent souvent de



convergences, et les caractères morphologiques partagés accessibles par microscopie optique sont trop peu nombreux chez ces organismes de petite taille pour déterminer les relations de parenté entre groupes ou même de classer tous les micro-organismes observables.

L'apparition du microscope électronique au milieu du XX<sup>ème</sup> siècle améliore la situation. En effet, le pouvoir de grossissement de ces appareils, bien supérieur à celui de la microscopie optique, donne accès à de nombreux caractères ultrastructuraux (e.g. forme des replis de la membrane interne mitochondriale), dont le partage entre différentes espèces permet de définir des taxons considérés monophylétiques. Corliss en décompte 45 en 1984, dont de nombreux toujours d'actualité aujourd'hui tels que les 3 premiers grands groupes d'alvéolés décrits (ciliés, apicomplexés et dinoflagellés), les diatomées, et les chlorophytes (Corliss 1984). L'accès à un nouveau type de caractères en cette fin de XX<sup>ème</sup> siècle va de nouveau améliorer nos connaissances de la taxonomie des eucaryotes, et surtout permettre de déterminer les relations de parentés entre les groupes (Philippe *et al.* 2000).

#### 4) L'apport des données moléculaires.

##### *Les débuts de "l'ère moléculaire".*

En 1965, Zuckerkandl and Pauling proposent de retracer l'histoire évolutive des organismes en se basant sur des caractéristiques moléculaires plutôt que morphologiques ou physiologiques. Ils distinguent trois types de molécules selon la quantité d'informations historiques qu'elles comportent, et expliquent que seuls les polymères macromoléculaires portant directement dans leur séquence l'information génétique, peuvent être utilisées de façon fiable pour déterminer les relations entre organismes. Parmi ces molécules qu'ils nomment sémantides, on trouve l'ADN (sémantides primaires), les ARN messagers (sémantides secondaires) et les polypeptides synthétisés directement à partir d'un ARNm sans modification post-traductionnelle (sémantides tertiaires). Bien que leur étude ne soit pas encore techniquement possible, ils supposent que les sémantides primaires contiennent plus d'information du fait de la présence de séquences non transcrites et des mutations isosémantiques rendues invisibles dans la séquence des protéines.

Woese, Fox et collègues mettent en pratique cette idée, en comparant différents organismes et chloroplastes sur la base de la séquence de leurs ARNr 16S ou 18S. Leurs travaux font l'objet de plusieurs publications au cours de l'année 1977. A cette date, on ne peut séquencer que de tous petits fragments. Aussi, ils digèrent chaque ARN avec la RNase T1 (une endonucléase coupant spécifiquement la liaison phosphodiester en 3' d'un nucléotide à guanine). Après séparation par électrophorèse et séquençage des fragments ainsi obtenus, ils



## Introduction

obtiennent pour chaque organisme un catalogue d'oligonucléotides (Woese & Fox 1977). Ils comparent alors les catalogues deux à deux en calculant un coefficient  $S_{AB}$  correspondant à deux fois le nombre de nucléotides présents dans les oligonucléotides de plus de 5 nucléotides et communs aux deux catalogues divisé par le nombre de nucléotides présents dans l'ensemble des oligonucléotides de plus de 5 nucléotides des deux catalogues. Ils en déduisent une matrice de distance des organismes considérés, et mettent en évidence une division primaire en trois grands domaines, tous aussi distants les uns des autres, qu'ils nomment Eubacteria, Urkaryotes et Archaeobacteria (Woese & Fox 1977) puis Bacteria, Eucarya et Archaea (Woese *et al.* 1990). Ces travaux constituent une révolution dans les classifications de l'époque, balayant la classification en 5 règnes de Whittaker qui prévalait à l'époque ainsi que la distinction eucaryotes / procaryotes. Ils démontrent de plus la faisabilité de l'idée théorique de Zuckerkandl and Pauling (1965), et proposent pour la première fois l'utilisation du marqueur moléculaire encore le plus utilisé aujourd'hui, le gène codant l'ARN de la petite sous-unité du ribosome (16S ou 18S).

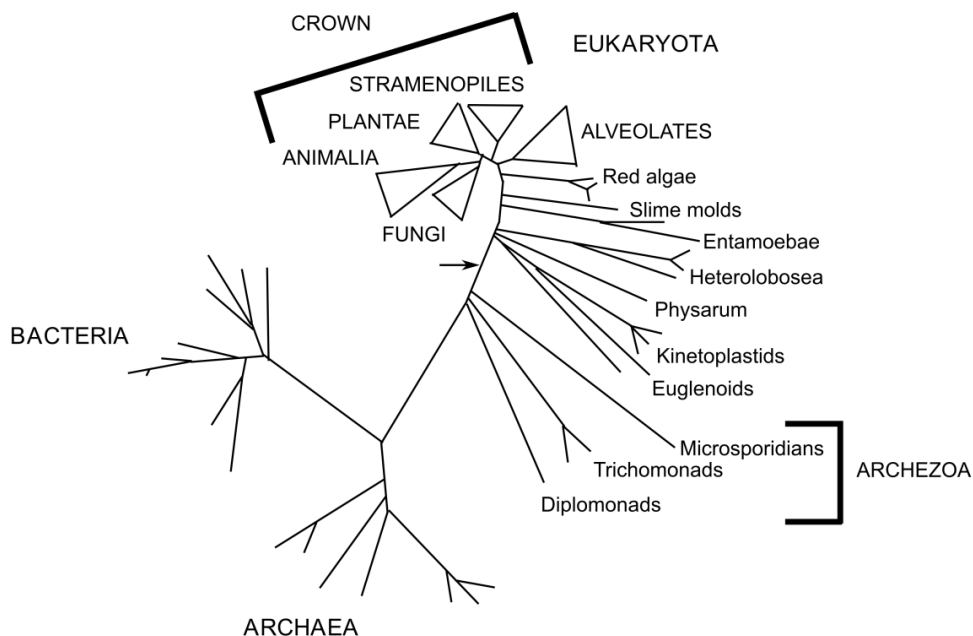
Quelques années plus tard, en 1986, Sogin et ses collaborateurs utilisent à leur tour la comparaison des séquences d'un gène, pour étudier spécifiquement les relations de parenté entre micro-organismes eucaryotes. Ils utilisent le gène codant l'ARN 18S de la petite sous-unité du ribosome. Ils comparent les séquences des ADNr 18S de plusieurs macro-organismes (le riz *Oryza sativa*, le maïs *Zea mays*, le rat *Ratus ratus*, le xénope *Xenopus laevis*, l'artémie *Artemisia salina*), de divers eucaryotes unicellulaires (e.g. les amibes *Acanthamoeba castellanii* et *Dictyostelium discoideum*, la levure *Saccharomyces cerevisiae*, les ciliés *Paramecium tetraurelia* et *Tetrahymena thermophila*, le trypanosome *Trypanosoma brucei*, et l'euglène *Euglena gracilis*), ainsi que les ADNr 16S de deux bactéries et deux archées. Ils exposent alors la polyphilie du règne Plantae puisque les euglènes qui y étaient classées du fait de leurs plastes semblent n'avoir qu'une lointaine parenté avec les plantes terrestres, et surtout celle du règne des protistes dans lequel la diversité semble plus grande que dans les trois autres règnes eucaryotes réunis (végétaux, animaux et champignons).

Avec l'arrivée de ces méthodes moléculaires, les microbiologistes ne sont plus limités par le peu de caractères phénotypiques observables, et la détermination d'une classification représentant au plus juste les relations de parenté entre eucaryotes unicellulaires paraît plus réalisable.

## Des premiers arbres « 18S » aux classifications actuelles.

A début des années 1990, les premières phylogénies moléculaires sont basées sur les séquences des gènes de l'ARNr de la petite sous-unité du ribosome (ARNr 18S dans le cas des

eucaryotes). Ces premières phylogénies séparent les eucaryotes en un nombre de grands groupes relativement élevé. Plantes terrestres, animaux, champignons, straménopiles, alvéolés et algues rouges forment une « couronne » au sein de laquelle les relations ne sont pas définies. A la base de cette « couronne » émergent en échelle un certain nombre de groupes tels que les euglènes, des amibes, et les trois premiers groupes à diverger : microsporidies (ex : *Encephalitozoon* sp.), diplomonadines (ex : *Giardia lamblia*) et trichomonadines (ex : *Trichomonas vaginalis*) (Figure 3). Dès le début des années 1980, Cavalier-Smith propose que microsporidies, trichomonades et diplomonades, tous amitochondriaux, forment un sous-règne au sein des Protozoa, sous-règne qui aurait divergé avant l'endosymbiose mitochondriale. La position basale de ces lignées dans les premières phylogénies moléculaires des eucaryotes semble conforter cette « hypothèse Archezoa ».



**Figure 3.** *Arbre du vivant basé sur l'ARN de la petite sous-unité du ribosome. La flèche indique l'acquisition supposée de la mitochondrie dans l'arbre des eucaryotes. Modifié d'après (Katz 1998)*

Cependant, cette hypothèse est bientôt mise à mal. En effet, les preuves indiquant l'existence d'ancêtres des lignées Archezoa pourvus de mitochondries s'accumulent. Des homologues de gènes typiquement mitochondriaux chez les animaux sont découverts dans les génomes de ces organismes (e.g. découverte du gène codant la protéine chaperonne Hsp60 chez *Giardia lamblia*, Soltys and Gupta, 1994). On observe également chez de nombreux organismes amitochondriaux des organites membranaires, les mitosomes (découverts chez

## Introduction

*Entamoeba histolytica* en 1999, Tovar et al., 1999) et hydrogénosomes (découverts chez *Tritrichomonas fetus* en 1973, Lindmark and Müller, 1973), qui apparaissent dérivés des mitochondries classiques. Ainsi les lignées Archezoa ne semblent pas être primitivement amitochondriales mais au contraire particulièrement dérivées. Philippe et collaborateurs (2000) proposent alors que ces lignées se retrouvent artefactuellement à la base des eucaryotes suite à un phénomène d'attraction de longue branche. Cette attraction serait due notamment à la saturation mutationnelle de l'ADNr 18S : une même position peut être affectée par plusieurs événements de substitution successifs, le signal phylogénétique le plus ancien est progressivement effacé et le gène marqueur n'est plus informatif.

Par la suite, l'augmentation de l'échantillonnage taxonomique, l'utilisation de gènes autres que l'ADNr 18s (gènes codant les chaperonnes Hsp70 et Hsp90, ou les monomères de microtubules  $\alpha$  et  $\beta$ -tubulines notamment), seuls ou en concaténations, ainsi que de meilleurs modèles d'évolution (prenant en compte les différences de vitesses d'évolution entre sites d'une séquence) remodelent totalement la phylogénie des eucaryotes. Les branches basales se retrouvent alors dispersées au sein des groupes de la « couronne », e.g. les microsporidies parmi les champignons, les diplomonades dans les excavés. Le nombre de grands groupes d'eucaryotes diminue fortement par rapport aux premières phylogénies moléculaires, et leur monophylie semble bien soutenue (Philippe *et al.* 2000). Les relations entre ces grands groupes restent cependant mal résolues.

On considère aujourd'hui que les eucaryotes se regroupent en sept supergroupes, les opisthocontes, amibes, excavés, straménopiles, alvéolés, rhizariés, et archaeplastides (Plantae), ainsi que quelques groupes à l'affiliation incertaine.

## B- La diversité des eucaryotes unicellulaires à l'heure actuelle.

### 1) Décrire la diversité aujourd'hui.

#### *La nécessité des méthodes moléculaires.*

Les méthodes classiques d'identification des micro-organismes se basent sur des observations morphologiques en microscopie et des tests métaboliques réalisés sur des organismes maintenus en culture. Or, comme nous l'avons vu, les critères morphologiques et fonctionnels observables sont souvent trop limités pour permettre l'affiliation précise d'un organisme microscopique à un taxon. Certains organismes morphologiquement très semblables peuvent être phylogénétiquement très différents. Ainsi, les phylogénies moléculaires ont révélé que certains groupes d'organismes considérés sur la base de leur morphologie comme une seule et même espèce regroupaient en fait des lignées très

différentes, baptisées espèces cryptiques (Šlapeta *et al.* 2006). En outre, on n'arrive à maintenir en culture qu'une toute petite partie des micro-organismes (Zinger *et al.* 2011), qui ne correspond pas nécessairement aux espèces les plus abondantes (Johnson & Sieburth 1982; Moreira & López-García 2002; Massana *et al.* 2004a).

### ***Quelques méthodes moléculaires.***

Il ne s'agit pas ici de dresser une liste exhaustive de toutes les méthodes moléculaires mises en œuvre pour étudier la diversité des micro-organismes eucaryotes. Nous tâcherons de nous concentrer sur les méthodes qui sont ou ont été les plus utilisées.

La plupart des méthodes que nous présenterons se basent sur l'amplification de marqueurs moléculaires, directement à partir d'ADN extrait d'un échantillon environnemental, sans étape de mise en culture. Dans la pratique, l'ensemble de l'ADN contenu dans un échantillon est extrait, puis un marqueur moléculaire est amplifié par réaction de polymérisation en chaîne (PCR). Les marqueurs moléculaires sont des régions du génome conservées chez l'ensemble des organismes considérés, dont la séquence, telle un code-barres, permet d'identifier le groupe auquel appartient l'organisme. Les gènes codant les ARN ribosomiques sont les plus utilisés. Notamment, le gène codant l'ARN de la petite sous-unité du ribosome (ADNr 18S chez les eucaryotes) est le plus souvent ciblé. De fait, la richesse des bibliothèques de séquences qui en découle rend ce marqueur de plus en plus informatif. L'ADNr 28S, plus long, ou l'ADNr 16S chloroplastidial dans le cas d'eucaryotes photosynthétiques sont également utilisés. Ces marqueurs sont amplifiés à l'aide d'amorces générales des eucaryotes, ou bien ciblant des taxa d'intérêt, notamment lorsque ceux-ci sont peu abondants, ou que leurs gènes sont mal amplifiés par les amorces générales. A partir du pool d'amplicons obtenu, plusieurs méthodes non mutuellement exclusives peuvent être mises en œuvre pour étudier la diversité moléculaire présente dans le pool d'amplicons et par extension dans l'échantillon environnemental.

### ***→Quantifier la diversité, comparer des échantillons.***

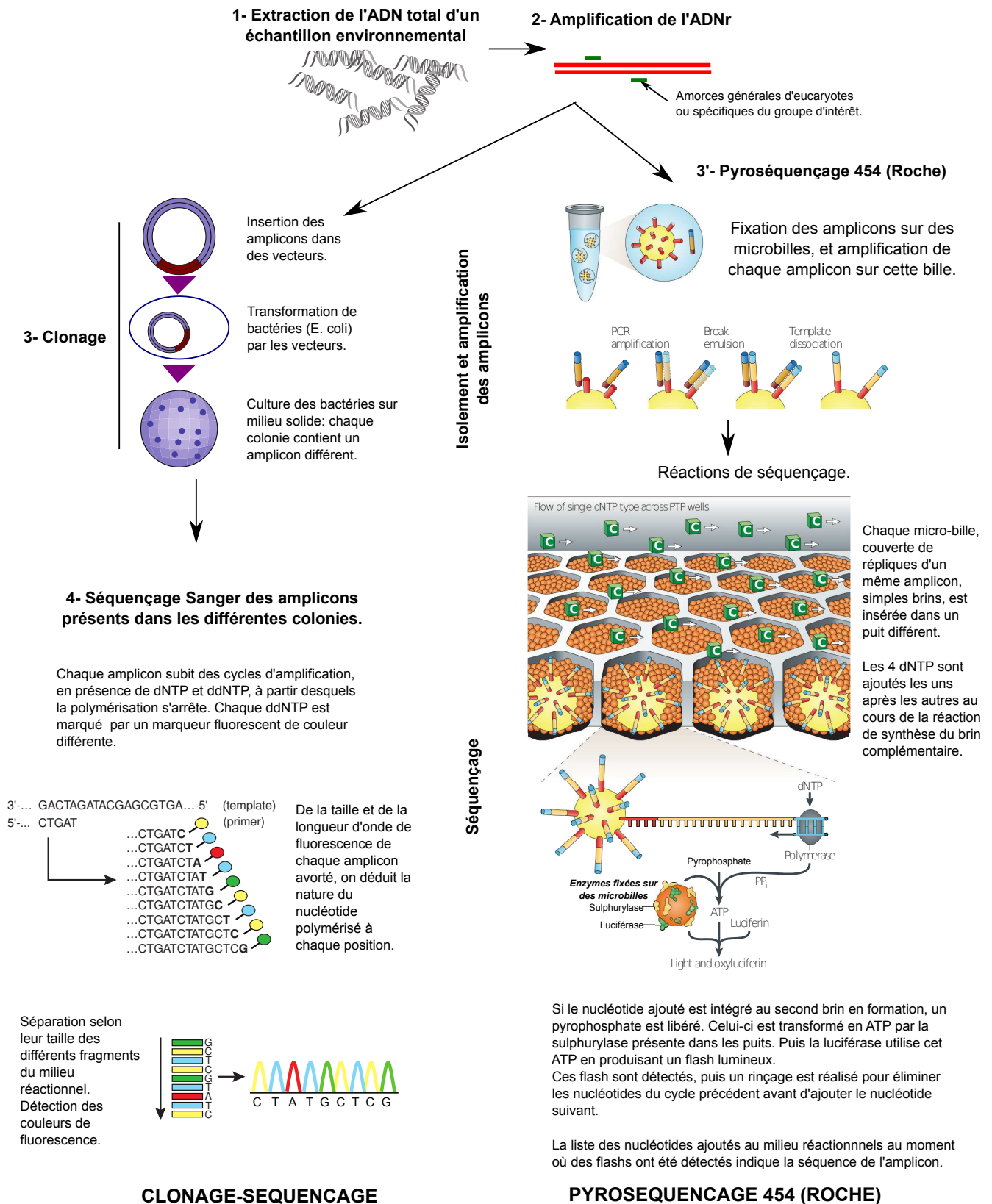
Tout d'abord, des techniques « d'empreinte moléculaire » peuvent être utilisées. Ces méthodes ne permettent pas directement d'identifier les organismes présents dans l'échantillon environnemental, mais peuvent permettre de comparer semi-quantitativement la diversité présente dans différents échantillons, ou d'estimer la part de la diversité moléculaire partagée par ces échantillons. L'électrophorèse sur gel avec gradient de dénaturation (DGGE) et le polymorphisme de longueur du fragment de restriction terminal (T-RFLP) sont sans doute les deux méthodes d'empreinte moléculaire les plus courantes. Au cours de la DGGE,

## *Introduction*

les amplicons obtenus par PCR sont séparés par électrophorèse en fonction de leur propension à être facilement dénaturés, c'est-à-dire en fonction de leur séquence et plus particulièrement de leur teneur en GC. Le T-RFLP est basé sur la digestion par des enzymes de restriction des amplicons marqués par fluorescence à l'une des extrémités. Après migration électrophorétique des fragments obtenus et révélation de l'extrémité marquée, les différentes bandes correspondent à des amplicons de séquences différentes. Dans les deux cas, comparer le nombre et la position des bandes sur le gel d'électrophorèse pour différents échantillons permet de comparer la diversité présente dans ces échantillons. Ces méthodes peuvent ensuite être couplées à un séquençage pour identifier les organismes correspondant à chaque bande.

→ *Identifier les micro-organismes présents : déterminer la composition des communautés.*

Les amplicons issus de la PCR sur l'ADN total issu des échantillons environnementaux peuvent également être séquencés, par plusieurs méthodes, notamment le clonage et séquençage Sanger et, plus récemment, le séquençage direct des amplicons par pyroséquençage ou séquençage Illumina (Figure 4). Les séquences ainsi obtenues sont supposées représentatives de la diversité génétique des amplicons, et donc de celle des organismes présents dans l'échantillon de départ. Les séquences partageant plus d'un certain pourcentage d'identité (en général 98% chez les eucaryotes; Caron and Countway, 2009) sont groupées en unités taxonomiques opérationnelles (OTU), considérées comme un proxy des espèces. Les séquences représentatives de chaque OTU sont comparées à des séquences d'organismes identifiés (par analyses BLAST et/ou reconstructions phylogénétiques). Elles sont ainsi affiliées au même groupe taxonomique que les séquences de référence proches, si elles existent. Certains groupes d'OTU environnementales ne sont proches d'aucune séquence d'organismes cultivés, et représentent une part encore non décrite de la diversité microbienne (López-García & Moreira 2008). Les proportions de chaque OTU au sein de l'ensemble des séquences issues de l'échantillon représentent de façon semi-quantitative l'abondance relative de l'espèce correspondante dans cet échantillon.



**Figure 4: Principales étapes et principes de l'obtention de séquences environnementales par les méthodes de clonage-séquençage Sanger et pyroséquençage 454. Illustrations empruntées à (Shendure & Ji 2008; Metzker 2009).**



## Introduction

### → *Distribution et abondance relative.*

La distribution et l'abondance d'un groupe donné peuvent également être déterminées sans passer par une étape de séquençage. Par exemple, la technique d'hybridation *in situ* (FISH) permet de marquer spécifiquement les cellules d'un groupe d'intérêt, à l'aide d'une sonde fluorescente ciblant les gènes de l'ADNr 18S ou 16S d'un groupe donné. Par la suite, un comptage manuel sous microscope permet de déterminer l'abondance relative du groupe cible, parmi l'ensemble des cellules de l'échantillon. Le FISH peut également permettre d'observer des cellules dont on ne connaît que la séquence d'un marqueur moléculaire (Kolodziej & Stoeck 2007; Jones *et al.* 2011). D'autre part, certains groupes d'organismes photosynthétiques contiennent des pigments qui leur sont propres. Doser ces pigments peut permettre d'estimer l'abondance des groupes en questions. Par exemple, le 19'-hexanoyloxyfuxoxanthine (19-Hex) est utilisé pour montrer la prédominance des haptophytes dans le phytoplancton marin (Andersen 2004; Liu *et al.* 2009).

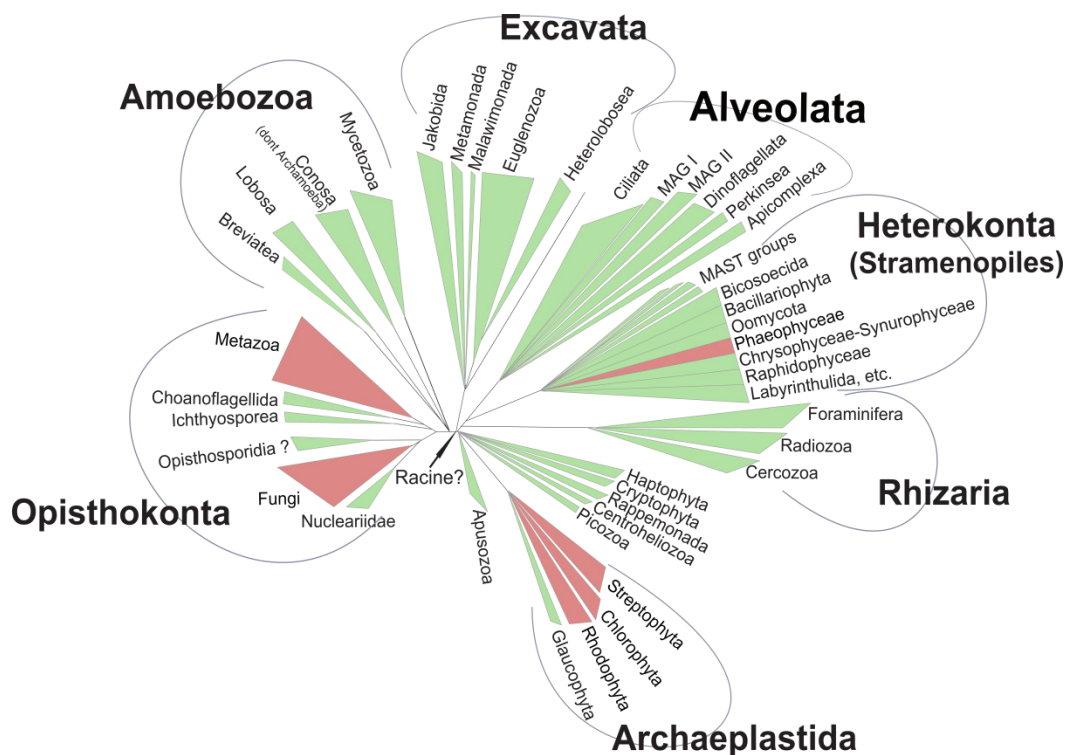
### ***Limites des méthodes moléculaires et importance de l'observation directe des organismes.***

Rapidement, les méthodes moléculaires permettant de s'affranchir de l'étape de mise en culture ont été utilisées pour explorer la diversité bactérienne dans tous types d'environnement. Cette approche n'a été mise en œuvre pour étudier la diversité eucaryote que quelques années plus tard, mais a également permis de découvrir un large pan de la diversité sans représentants cultivés (Moreira & López-García 2002). Cependant, les méthodes moléculaires ont des biais. Qualitativement d'abord, une partie des OTU considérées comme représentant des lignées divergentes pourraient en fait être artefactuelles (e.g. chimères, erreurs de séquençage (Kunin *et al.* 2010)). Quantitativement ensuite, l'ADN n'est pas aussi bien extrait de toutes les cellules, les amorces, surtout générales, ne permettent pas d'amplifier aussi bien les marqueurs moléculaires de toutes les espèces (Zinger *et al.* 2011), et le nombre de copies d'ADNr 18S dans les génomes peut fortement varier selon les taxa. Ainsi, la proportion de séquences affiliées à un groupe donné dans une librairie de séquences ne correspond pas toujours exactement à la proportion de ce groupe dans l'échantillon environnemental initial. Enfin, l'approche moléculaire ne peut en aucun cas se substituer totalement à la description classique des espèces. En effet, l'affiliation des séquences environnementales à un taxon n'est possible que s'il existe dans les bases de données des séquences associées à des espèces décrites. Ensuite, les méthodes moléculaires ne donnent aucune information précise sur la biologie, le métabolisme, ou l'écologie des organismes détectés. Comme évoqué plus haut, le FISH peut permettre d'observer certains organismes connus uniquement par une séquence, mais cette méthode a des biais, notamment

due à la fixation des cellules, et les observations réalisées ne sont pas toujours fiables. Par exemple, Not et collègues ont proposé en 2007 qu'un groupe de séquences environnementales puisse correspondre à des organismes photosynthétiques, qu'ils ont nommés "Picobiliphytes", en se basant sur des observations réalisées grâce au FISH. Par la suite, des études génomiques et surtout la mise en culture d'un représentant de ce groupe ont montré qu'il s'agissait en réalité d'un groupe d'organismes hétérotrophes, bactériovores, rebaptisés alors "Picozoa" (Seenivasan *et al.* 2013; Moreira & López-García 2014). Ainsi, approches moléculaires et traditionnelles sont complémentaires.

## 2) La diversité phylogénétique eucaryote telle qu'on la connaît actuellement.

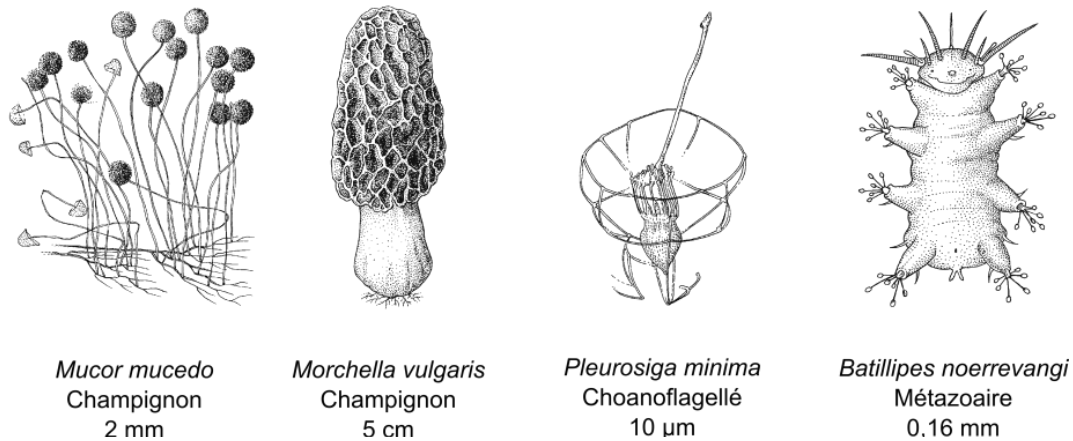
L'association des méthodes traditionnelles et moléculaires a permis de classer l'ensemble de la diversité eucaryote en sept grands supergroupes, auxquels s'ajoutent quelques taxa dont la position dans l'arbre phylogénétique eucaryote reste mal définie (Figure 5). Quand aucune autre source n'est précisée, les informations données sur chaque groupe sont issues de Simpson and Roger, 2004; Lecointre and Le Guyader, 2006; Baldauf, 2008; Bertrand *et al.*, 2011; Walker, RG Dorrell, *et al.*, 2011.



**Figure 5 :** Arbre phylogénétique schématisé présentant les relations entre les principaux groupes d'eucaryotes, basé sur des données moléculaires. Les groupes en rouge contiennent des macro-organismes, les autres regroupent uniquement des micro-organismes. Modifié à partir de López-García and Moreira, 2008.



*Les opisthocontes (Opisthokonta).*

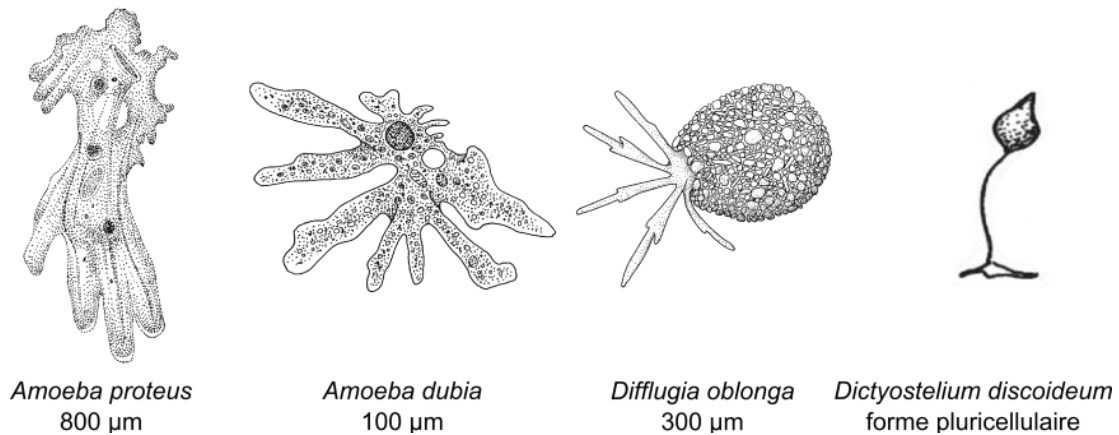


*Figure 6 : Quelques exemples d'opisthocontes (Lecointre & Le Guyader 2006).*

Les opisthocontes regroupent des organismes présentant une large diversité de morphologies, habitats et modes de vie. On y trouve deux des grands groupes contenant des organismes pluricellulaires : métazoaires (animaux) et champignons, ainsi que les choanoflagellés, les ichthyosporés et les nucléarides. Les métazoaires sont tous pluricellulaires et phagotrophes, prédateurs ou parasites. Les champignons se nourrissent par osmotrophie, ils sont essentiellement saprotrophes, parfois parasites ou symbiotes, et jouent ainsi un rôle crucial en tant que décomposeurs ou symbiotes d'embryophytes (e.g. mycorhizes), notamment dans les écosystèmes terrestres. Choanoflagellés et nucléarides sont des prédateurs libres, se déplaçant respectivement à l'aide d'un flagelle ou par déplacement amiboïde (via un pseudopode). Les ichthyosporés regroupent quant à eux des organismes parasites, essentiellement d'animaux. Récemment, un nouveau taxon a été proposé parmi les opisthocontes. Les Opisthosporidia (Karpov *et al.* 2014) regrouperaient les microsporidies, les Cryptomycota-Rozellida et les aphélides, tous parasites intracellulaires, osmotrophes ou phagotrophes.

Ce supergroupe a été formé au début des années 1990, sur la base de phylogénies moléculaires. La présence d'un flagelle unique et propulseur, à l'arrière de la cellule, est considérée comme une synapomorphie du groupe, ainsi que différents événements génétiques rares (e.g. insertion dans le gène du facteur d'élongation EF1- $\alpha$ ). Il semblerait que choanoflagellés, ichthyosporés et métazoaires d'une part, et champignons, nucléarides et opisthosporides d'autre part forment deux clades distincts

## Les Amoebozoa.



**Figure 7 : Quelques exemples d'amibes** (Lecointre & Le Guyader 2006; Walker *et al.* 2011a).

Le supergroupe Amoebozoa regroupe des organismes amiboïdes ou amiboflagellés. Ils se déplacent grâce à des pseudopodes tubulaires ou plats, ou, quand il est présent, à l'aide d'un flagelle unique situé à l'avant de la cellule. La plupart sont des organismes libres, vivant dans l'eau et le sol, où ils se nourrissent généralement par bactérivorie, engouffrant leurs proies à l'aide des pseudopodes. Certains taxons d'amibes, notamment les Archamoebae, sont parasites. *Entamoeba histolytica* est par exemple responsable de 40 000 à 100 000 décès par an, dans le monde entier, chez *Homo sapiens*. Dans ce taxon, les organismes sont anaérobies ou microaérophiles, et ont perdu secondairement leurs mitochondries, remplacés par des mitosomes. Certaines amibes (les amibes « sociales »), *Dictyostelium discoideum* notamment, forment par agrégation des structures pluricellulaires assurant la dispersion des spores, formes de résistance.

Le supergroupe Amoebozoa est relativement récent. Il résulte du regroupement par des phylogénies multigènes de taxa précédemment séparés dans l'arbre des eucaryotes. La phylogénie interne du groupe est encore incertaine. On suppose que les Amoebozoa seraient le groupe frère des Opisthoconta, formant à eux deux les unicontes (ils possèdent pour la plupart un flagelle unique).

On notera que la morphologie cellulaire 'amiboïde' n'est pas une synapomorphie des Amoebozoa. D'autres organismes appartenant à des supergroupes différents sont qualifiés d'amibes pour caractériser leur morphologie mais ne sont pas des Amoebozoa.

*Les excavés (Excavata).*

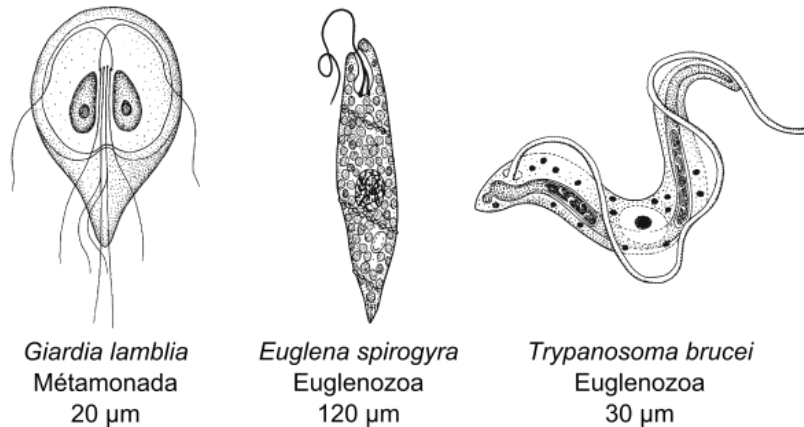


Figure 8 : Quelques exemples d'excavés (Lecointre & Le Guyader 2006).

Les excavés forment un groupe hétéroclite d'organismes unicellulaires. La plupart sont des flagellés hétérotrophes libres, vivant dans les sols ou en milieu aquatique. Ils se nourrissent de particules en suspension dirigées vers un sillon d'alimentation, à l'avant de la cellule, par un flagelle. D'autres cependant sont amiboïdes, comme les Heterolobosea. Il existe également des excavés commensaux, symbiontes ou pathogènes, mitochondriaux (e.g. *Trypanosoma brucei*) ou amitochondriaux (e.g. *Giardia intestinalis*). Enfin, au sein des euglenozoa, certains organismes sont photosynthétiques suite à une endosymbiose secondaire avec une algue verte.

La monophylie de ce groupe est souvent discutée, mais un ensemble de données moléculaires et morphologiques semble converger dans ce sens (e.g. structure du sillon d'alimentation, phylogénies multigènes à large échantillonnage taxonomique (Moreira et al., 2007)).

*Les alvéolés (Alveolata).*

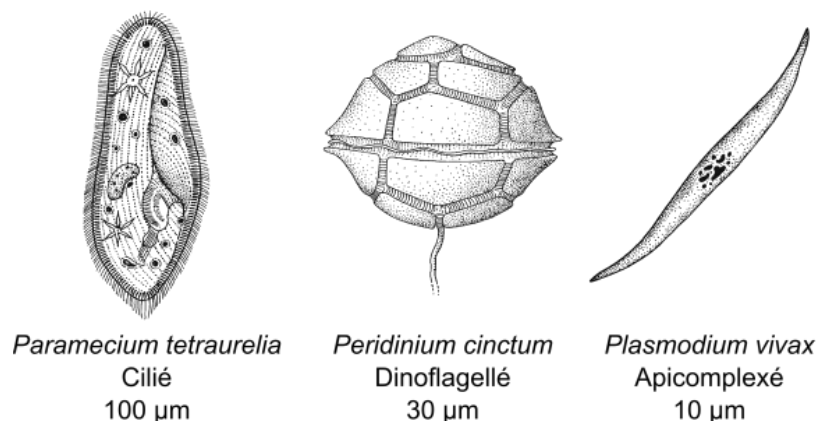
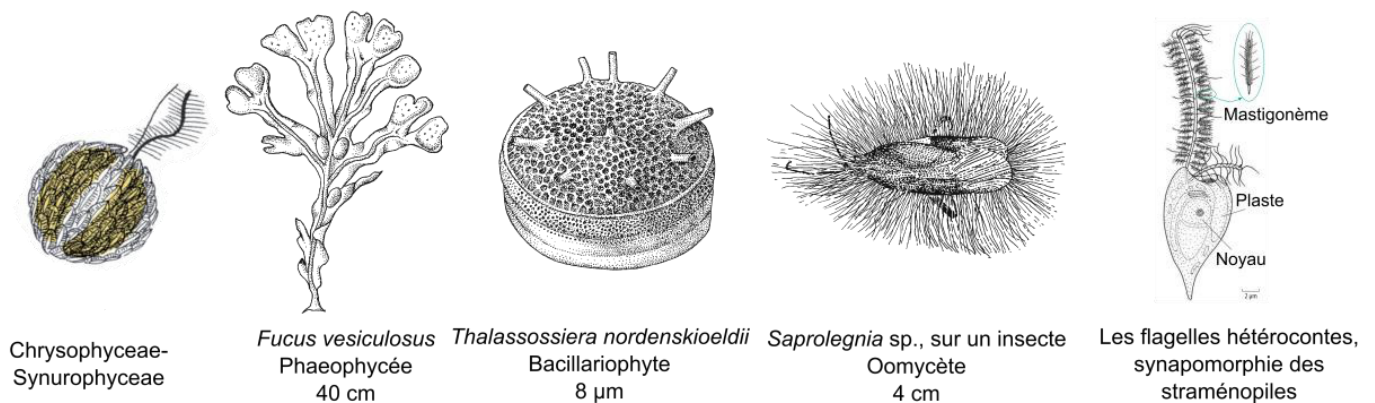


Figure 9 : Quelques exemples d'alvéolés (Lecointre & Le Guyader 2006).

Les alvéolés regroupent historiquement les ciliés, dinoflagellés et apicomplexés, auxquels se sont ajoutés de nouveaux taxa comme les perkinsozoa, les colpodellides ou les syndiniales. Tous possèdent, sous la membrane plasmique, une couche continue de vésicules : les alvéoles. Les ciliés sont hétérotrophes et prédateurs. Ils sont caractérisés par la présence de cils à leur surface, et la possession de deux types de noyaux, les micronoyaux impliqués dans la réplication du génome et un macronoyau dans lequel a lieu la transcription. Ils vivent en milieu aquatique, notamment dans les écosystèmes continentaux où ils peuvent être très abondants. Les apicomplexés sont des parasites intracellulaires obligatoires des métazoaires, le plus connu étant *Plasmodium falcifarum*. Ils sont caractérisés par la présence d'une structure apicale impliquée dans l'attachement et la pénétration dans la cellule hôte. Certains possèdent des apicoplastes, organites dérivés de chloroplastes mais ne réalisant plus la photosynthèse. Les colpodellides ont également une structure apicale impliquée dans l'attachement à d'autres cellules, mais qui est cette fois utilisée pour aspirer le cytoplasme de leurs proies. Il s'agit en effet de flagellés libres et prédateurs. Les dinoflagellés peuvent être photosynthétiques, mixotrophes ou hétérotrophes. Leur plaste est issu d'une endosymbiose secondaire avec une algue rouge, parfois suivi d'un remplacement additionnel par endosymbiose tertiaire avec une algue rouge ou verte. Ils se caractérisent par deux flagelles perpendiculaires, dont l'un est enroulé dans un sillon transversal. Ils peuvent être symbiotiques (e.g. zooxanthelles) ou libres. En milieu marin, ils peuvent être extrêmement abondants, jusqu'à former des efflorescences (marées rouges) à fort impact écologique, notamment par la production de neurotoxines. Les Perkinsea sont des organismes parasites, essentiellement de mollusques et de dinoflagellés. Ils peuvent être très abondants en eau douce (Mangot *et al.* 2010). Enfin, les syndiniales se sont révélés extrêmement abondants dans tous les environnements marins grâce aux méthodes moléculaires (ils correspondent aux groupes des séquences environnementales MALV I et II (López-García *et al.* 2001)). Ces parasites intracellulaires ont pour l'instant été rencontrés uniquement en milieu marin (Guillou *et al.* 2008).

La monophylie des trois grands groupes d'alvéolés (ciliés, dinophytes, apicomplexés) a été montrée très tôt par des phylogénies basées sur l'ADNr 18S (Gajadhar *et al.* 1991).

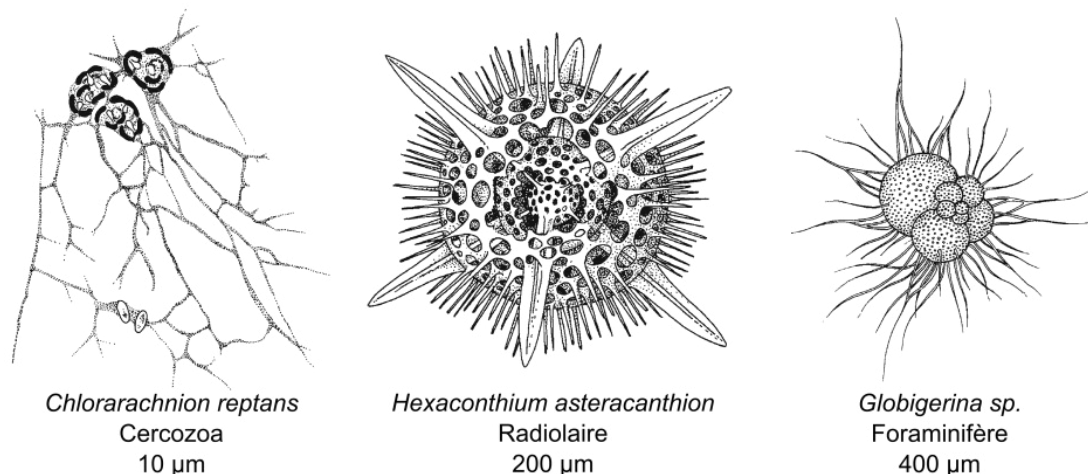
*Les straménopiles (Heterokonta).*



*Figure 10 : Quelques exemples de straménopiles (Lecointre & Le Guyader 2006; Walker et al. 2011a).*

Les straménopiles constituent le troisième supergroupe avec le plus grand nombre d'espèces décrites, après les opisthocontes et les archaeplastides (sans doute un biais vers les groupes contenant des macro-organismes). Ils sont caractérisés par la différence de taille entre leur deux flagelles, quand ils sont présents, et la présence sur le plus long de mastigonèmes (appendices protéiques) tubulaires et tripartites. Ils regroupent des taxa unicellulaires (e.g. bacillariophytes, chrysophytes) ou pluricellulaires (phaeophytes, xanthophytes). Certains sont photosynthétiques (e.g. phaeophytes, bacillariophytes, xanthophytes, chrysophytes). D'autres sont hétérotrophes. Les oomycètes par exemple sont d'importants parasites d'embryophytes (e.g. *Phytophthora infestans*), les labyrinthulomycètes sont saprotrophes, les bicosoécides prédateurs. Les straménopiles constituent d'importants producteurs primaires en milieu marin, et sont aussi abondants en eau douce (chrysophytes notamment) ou saumâtres (e.g. xanthophytes). Les straménopiles regroupent également un grand nombre de taxa presque totalement composés de séquences environnementales, sans représentants cultivés. Ces MAST (MARine STRamenopiles) sont ubiquitaires et très abondants en milieu marin, et certains ont été détectés en eau douce (Massana *et al.* 2013).

La monophylie du groupe se retrouve dans les caractéristiques des flagelles et dans les phylogénies multigènes.

*Les rhizariens (Rhizaria).*

*Figure 11 : Quelques exemples de rhizariens (Lecointre & Le Guyader 2006).*

Les Rhizaria constituent un groupe d'organismes divers. Ils se divisent en trois grands groupes. Les radiolaires sont des prédateurs exclusivement marins, planctoniques. Ils possèdent un squelette ("test") siliceux. Certains vivent en colonies. Les foraminifères sont amiboïdes, et possèdent des tests percés d'orifices par lesquels s'insèrent les pseudopodes. Ces tests sédimentent bien, et constituent des marqueurs biostratigraphiques et des témoins des paléotempératures très utilisés. Ils sont surtout abondants dans le benthos marin, mais on les rencontre également dans le plancton marin ou d'eau douce et les sols. Radiolaires et foraminifères peuvent former des colonies macroscopiques (Caron *et al.* 2012). Enfin, les cercozoaires constituent un groupe hétérogène d'organismes amiboïdes, flagellés ou amiboflagellés. La plupart sont hétérotrophes mais certains, les chlorarachniophytes sont mixotrophes. Ils possèdent un plaste issu d'une endosymbiose secondaire avec une algue verte, qui a conservé un nucléomorphe. Les chlorarachniophytes sont marins, les autres espèces de cercozoaires sont communes en eau douce ou dans les sols.

Bien qu'aucune synapomorphie n'ait pu être définie pour l'ensemble des Rhizaria, sa monophilie est bien soutenue par les phylogénies moléculaires (Moreira *et al.* 2007). En revanche les relations au sein de ce supergroupe sont mal définies. Les Rhizaria forment un groupe monophylétique ("SAR") avec les straménopiles et les alvéolés.



## Les archaeplastidés (Archaeplastida).

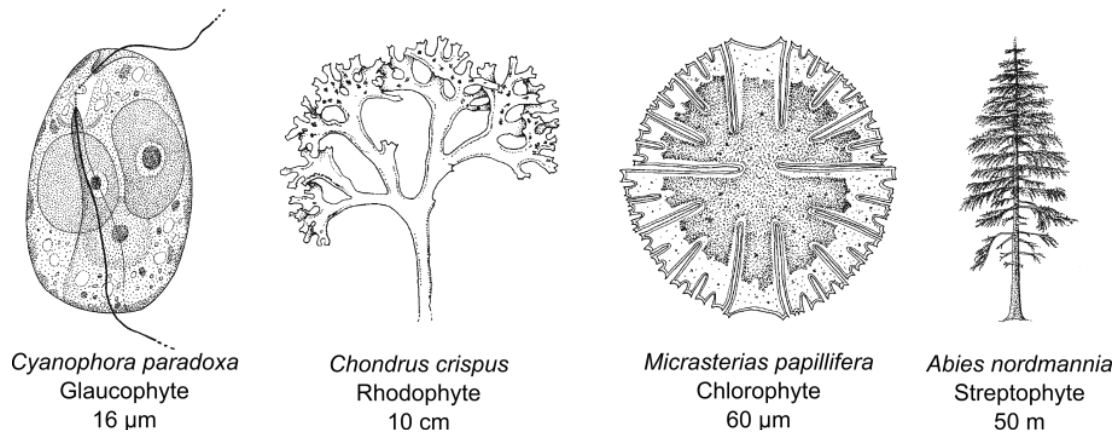


Figure 12 : Quelques exemples d'archaeplastidés (Lecointre & Le Guyader 2006).

C'est à la base de ce groupe qu'est apparue la photosynthèse eucaryote, par endosymbiose primaire avec une cyanobactérie. La quasi-totalité des Archaeplastida sont photosynthétiques ; ils ont colonisé tous les milieux, terrestres et aquatiques. On distingue trois groupes principaux, les Viridiplantae ou Chlorobiontes, les Rhodophytes et les Glaucophytes. Les Viridiplantae regroupent les streptophytes et les chlorophytes. Ils sont très abondants en milieu marin, d'eau douce, et dans les écosystèmes terrestres. Comme eux, les rhodophytes groupent des organismes uni et pluricellulaires. On suppose d'ailleurs plusieurs apparitions indépendantes de la pluricellularité chez les Archaeplastida. Les rhodophytes sont essentiellement marins, benthiques ou pélagiques, mais quelques espèces sont présentes en eau douce. Ils sont capables de réaliser la photosynthèse même à faible luminosité. Les glaucophytes sont généralement moins abondants, présents dans des écosystèmes d'eau douce et terrestres.

## Incertae sedis.

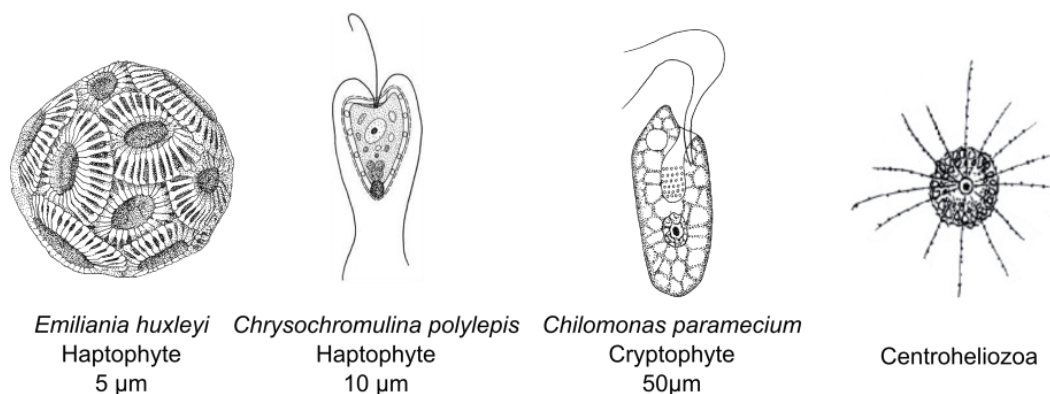


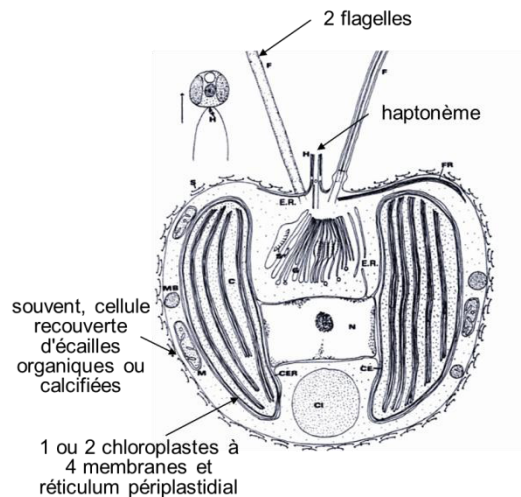
Figure 13 : Quelques exemples de taxa de position mal résolue dans l'arbre des eucaryotes (Lecointre & Le Guyader 2006; Walker et al. 2011a).

De nombreux taxa, plus ou moins abondants ou diversifiés ont encore une position mal résolue dans l'arbre des eucaryotes, et ne sont clairement affiliés à aucun des supergroupes présentés ci-dessus. Les Apusozoa regroupent des petits organismes unicellulaires, biflagellés. Les Katablepharida sont des flagellés libres, hétérotrophes, parfois possédant des kleptoplastes, c'est-à-dire des chloroplastes d'algues maintenus dans un prédateur après que celui-ci ait digéré le reste de l'algue. Ils vivent en milieu marin et en eau douce. Les Telonemida sont également des flagellés hétérotrophes, du plancton marin. Quant aux Centrohelioczoa, ils regroupent des organismes hétérotrophes amiboïdes, dont le corps cellulaire s'entoure de nombreux bras. Les Cryptophytes sont des membres abondants du plancton marin et d'eau douce. On les trouve également dans les eaux interstitielles des sols. Certains sont parasites, ou endosymbiontes de dinoflagellés. Il s'agit alors d'une endosymbiose tertiaire puisque les cryptophytes possèdent un plaste issu d'une endosymbiose secondaire avec un rhodophyte, plaste qui a conservé un nucléomorphe. Ils sont autotrophes ou mixotrophes pour la plupart, mais certaines espèces ont perdu secondairement leur plaste. Ils portent deux flagelles, couverts respectivement d'une et deux rangées de mastigonèmes, synapomorphie du groupe. Les haptophytes sont des membres essentiels du plancton océanique, dans toutes les régions du globe (Liu *et al.* 2009), mais leur diversité en eau douce est moins bien connue. Ils jouent un rôle important dans les réseaux trophiques, à la fois en tant que producteurs primaires et bactériophages. Ils peuvent atteindre des abondances extrêmement élevées, lors d'efflorescences (Edwardsen & Imai 2006). Certaines espèces produisent des toxines létales pour de nombreux organismes (Jordan & Chamberlain 1997). Ils peuvent être nus ou recouverts d'écailles (Figure 14), formées dans l'appareil de Golgi, organiques ou de calcaire (chez les coccolithophores) qui, en sédimentant, forment la craie. Ils sont impliqués dans le cycle du carbone, en fixant le CO<sub>2</sub> par photosynthèse et en le piégeant dans la craie (les coccolithophores seraient responsables de 25% du transport de carbone vers le fond de l'océan (Graham & Wilcox 2000)), et dans celui du soufre par la production de diméthyl sulfure (DMS), un composé actif sur le climat. Les haptophytes se divisent en deux grands groupes, les prymnésiophytes et les pavlovophytes, plus quelques groupes constitués pour l'instant uniquement de séquences environnementales (Šlapeta *et al.* 2005; Shi *et al.* 2009). Ils se caractérisent par la présence d'une structure particulière, l'haptonème, émergeant entre les deux flagelles et impliqué dans l'adhérence et la prédation. Ils possèdent un plaste à quatre membranes, issu d'une endosymbiose secondaire avec une algue rouge. A cause de ce plaste, ils ont longtemps été classés parmi les Chromalvéolés, avec les straménopiles, les alvéolés et les cryptophytes notamment, dont certains représentants possèdent également un plaste issu d'une endosymbiose avec une algue rouge (Keeling 2009). Ce groupe était soutenu



## Introduction

par des phylogénies basées sur des gènes plastidiaux, mais a été montré polyphylétique par des analyses basées sur des gènes nucléaires. Une hypothèse plus récente les place dans un taxon Hacrobia (ou CCTH), avec les cryptophytes, les centrohéliozoaires et les télonémides (Okamoto *et al.* 2009), mais la monophylie de ce groupe est également discutée (Pawlowski 2013).



**Figure 14 : Structure cellulaire d'un haptophyte** (Dessin : Chrysochromulina parva. Modifié à partir de Lee, 2008).

Contrairement aux bactéries et archées, les études de diversité basées sur des méthodes moléculaires n'ont pas mis en évidence de nouveau supergroupe clair (López-García & Moreira 2008; Massana & Pedrós-Alió 2008). Cependant, chaque nouvelle exploration moléculaire de la diversité eucaryote met en évidence l'existence d'une grande part de diversité encore inconnue au sein des groupes déjà connus présentés ci-dessus. En effet, une grande partie des phylotypes environnementaux diffère des séquences d'espèces connues, et même des séquences environnementales issues d'études précédentes (Shi *et al.* 2009; Triadó-Margarit & Casamayor 2012), suggérant qu'une partie de la diversité eucaryote reste encore à explorer. Au sein des supergroupes existant, quelques groupes de rang taxonomique relativement élevé ont même été observés, certains très abondants et répandus en milieu marin. Pourtant, ils restent connus essentiellement par des séquences environnementales, et ne sont représentées que par très peu d'organismes cultivés. C'est le cas par exemple des syndiniales (= MALV I et II) (Diez *et al.* 2001; Moon-van der Staay *et al.* 2001; López-García *et al.* 2001; Guillou *et al.* 2008) et des groupes de MAST (Massana *et al.* 2013) dont très peu ont des représentants décrits (Gómez *et al.* 2011).

### 3) Diversité métabolique et écologique des eucaryotes unicellulaires.

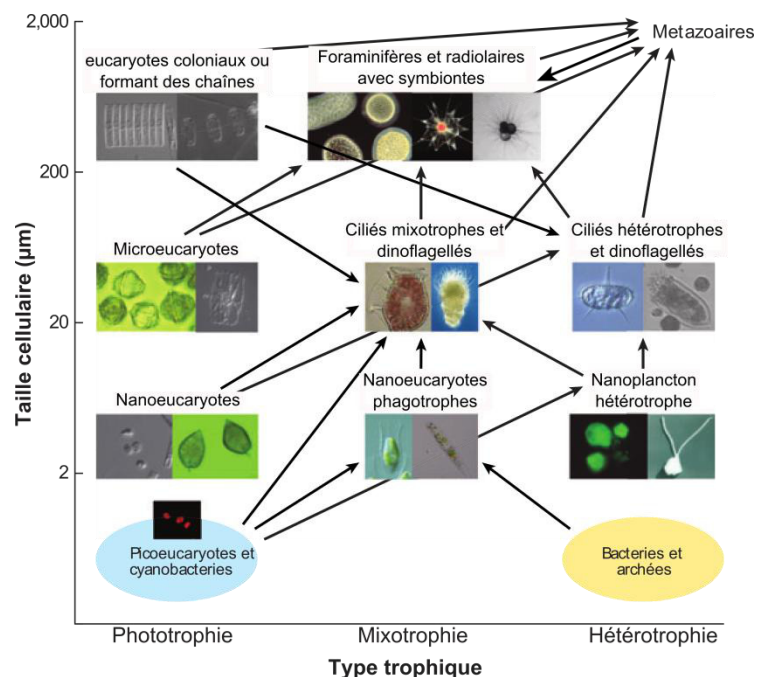
Comme on l'aura noté dans cette rapide présentation des principaux groupes, les eucaryotes unicellulaires présentent une grande diversité de types trophiques (phototrophie,

hétérotrophie et mixotrophie) et de modes de vie (libres, parasites ou symbiotes mutualistes). Ils ont colonisé l'ensemble des habitats marins, d'eaux douces ou terrestres. On les retrouve même dans des environnements considérés extrêmes comme les grandes profondeurs océaniques (López-García *et al.* 2001), les sources hydrothermales sous-marines (López-García *et al.* 2003), les environnements hyper-acides (Amaral Zettler *et al.* 2002) ou la glace (Bachy *et al.* 2011; Piwosz *et al.* 2013). Ils sont également capables d'entraîner la précipitation de différents minéraux : des carbonates de calcium (coccolithophores, foraminifères) ou de la silice (diatomées, radiolaires). Ils produisent une diversité de toxines agissant sur différents organismes (e.g. neurotoxines de dinoflagellés toxiques pour les métazoaires, acide acrylique bactéricide de certains haptophytes).

En raison de leur abondance et de la diversité des fonctions qu'ils réalisent, les micro-organismes eucaryotes jouent des rôles fondamentaux à la fois au niveau de leur écosystème (on les retrouve à tous les niveaux des réseaux trophiques Figure 15) et à l'échelle globale, par leur participation aux cycles du carbone (qu'ils piègent dans la matière organique ou en précipitant des carbonates) et du soufre. Ils ont notamment un impact non négligeable sur le climat, par exemple via leur production de diméthyl-sulfoniopropionate (DMSP) transformé par les bactéries en diméthylsulfide (DMS) qui, libéré dans l'atmosphère, entraîne la nucléation des nuages, et de là augmente l'albédo terrestre (Jordan & Chamberlain 1997; Simó 2001).

**Figure 15: Les micro-organismes eucaryotes dans un réseau trophique marin planctonique.**

On considère que les rôles joués par les organismes dépendent de leur taille (les prédateurs mangent en général des proies plus petites qu'eux) et de leur type trophique. Les flèches indiquent le sens et la direction des flux de matière organique. Figure empruntée à Caron *et al.*, 2012.



## C- Les petits eucaryotes.

Les micro-organismes eucaryotes présentent également une très grande diversité de taille, depuis *Ostrococcus tauri*, le plus petit eucaryote connu, qui mesure moins de 1  $\mu\text{m}$  de diamètre (Derelle *et al.* 2006) jusqu'à des organismes pouvant mesurer plusieurs millimètres de diamètre (e.g. le foraminifère *Hastigerina pelagica*, 10 000  $\mu\text{m}$  (Caron *et al.* 2012)). Les micro-organismes eucaryotes planctoniques sont classés en différentes fractions de taille : le picoplancton (0,2 à 2  $\mu\text{m}$ ), le nanoplancton (2 à 20  $\mu\text{m}$ ) et le microplancton (20 à 200  $\mu\text{m}$ ) (Bertrand *et al.* 2011). Les petits eucaryotes (pico et nanoplanctoniques) des milieux aquatiques suscitent depuis déjà quelques années un engouement particulier.

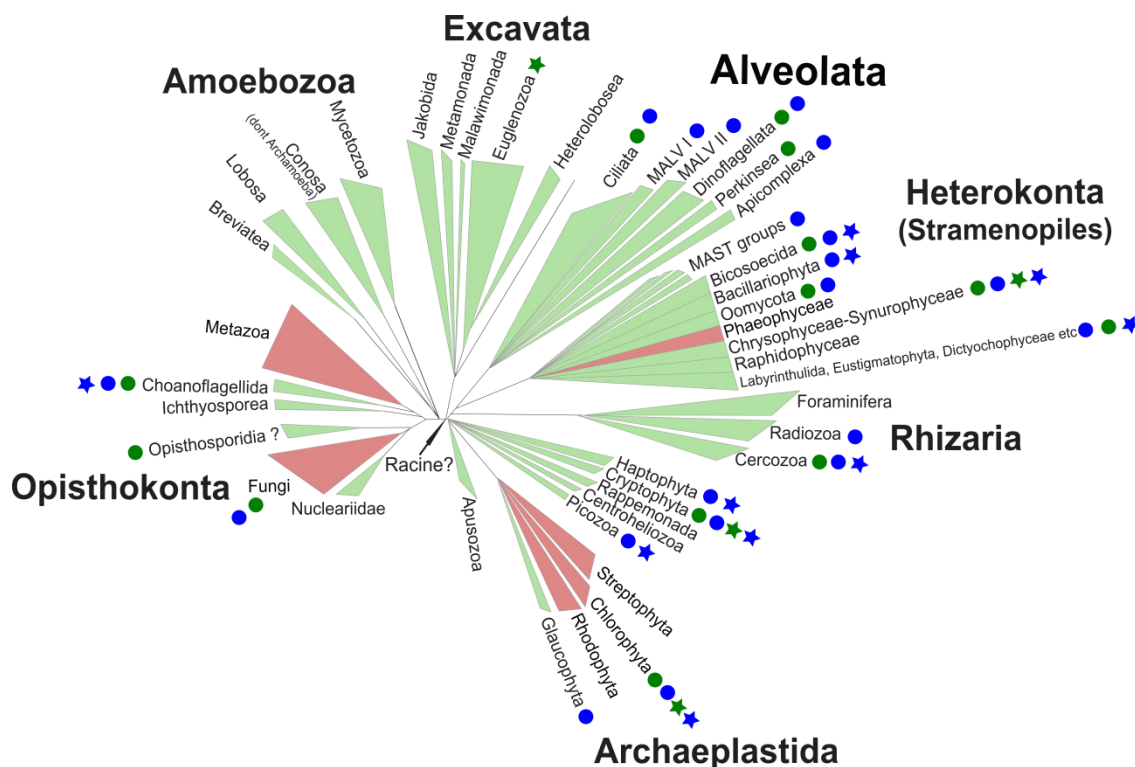
### 1) Diversité.

L'existence d'eucaryotes de taille similaire à celle de la plupart des bactéries est connue depuis maintenant plusieurs décennies. En 1952, Butcher applique aux eucaryotes une méthode d'isolement en culture traditionnellement utilisée pour les bactéries, par dilutions sériées, et isole pour la première fois des pico et nanoeucaryotes. A partir d'échantillons marins côtiers, il obtient quinze cultures de nanoeucaryotes planctoniques, parmi lesquelles sont des organismes de très petite taille, par exemple *Chromulina pusilla* (Chrysophyte) de 1-1,5  $\mu\text{m}$  de diamètre. Par la suite, différentes espèces de pico et nanoeucaryotes ont été décrites, la plupart parmi les espèces phototrophes. Notamment, les chlorophytes (surtout prasinophytes) et cryptophytes sont souvent observés (Johnson & Sieburth 1982; Stockner 1988), beaucoup appartiennent à de nouvelles classes comme les Bolidophyceae (Straménopiles) (Guillou *et al.* 1999). Certains organismes hétérotrophes sont également décrits (Fenchel 1982), mais beaucoup moins fréquemment, peut-être à cause de leur fragilité face aux méthodes de fixation ou des méthodes de culture peu adaptées (Pernthaler 2005). Cependant, de par leur petite taille, les pico et nanoeucaryotes présentent particulièrement peu de caractéristiques morphologiques distinctives, et leur taxonomie est mal connue (Stockner 1988).

Du moins jusqu'à l'arrivée des méthodes moléculaires d'exploration de la diversité. Les premières études utilisant ces méthodes pour les eucaryotes ciblaient en effet spécifiquement ceux de petite taille. La toute première a été réalisée en 1998, accidentellement, alors que les auteurs examinaient la diversité du bactérioplancton en milieu marin côtier (Rappé *et al.* 1998). Au cours de cette analyse, Rappé et collègues amplifient une forte proportion de gènes d'ARNr 16S d'origine plastidiale, grâce auxquels ils détectent la présence d'un grand nombre

de taxa eucaryotes photosynthétiques dans la fraction de taille 0,2-10  $\mu\text{m}$ , au sein des Prymnesiophyceae (Haptophytes), Bacillariophyceae (Straménopiles), Cryptophyceae et Prasinophyceae (Archaeplastidia). Par la suite, trois études, basées sur la diversité des ADNr 18S et ciblant les petits eucaryotes planctoniques dans différentes régions marines du globe, sont publiées en 2001 (Diez *et al.* 2001; Moon-van der Staay *et al.* 2001; López-García *et al.* 2001). Elles révèlent toutes trois une diversité bien supérieure à ce qui était suspecté, au sein d'une grande variété de taxa autotrophes et hétérotrophes, dans tous les supergroupes connus.

A l'heure actuelle, la présence de petits eucaryotes a été mise en évidence au sein tous les supergroupes (Figure 16), et souvent dans une grande diversité de taxa au sein de ces supergroupes.



**Figure 16: Les petits eucaryotes dans l'arbre du domaine.** Les disques indiquent la détection du groupe considéré par des méthodes moléculaires (basées sur la séquence des ADNr18S) dans les études suivantes: Diez *et al.*, 2001; López-García *et al.*, 2001; Moon-van der Staay *et al.*, 2001; Massana *et al.*, 2004; Romari and Vaulot, 2004; Richards and Bass, 2005; Šlapeta *et al.*, 2005; Lepère and Boucher, 2006; Not and Gausling, 2007; Lefèvre *et al.*, 2008 réalisées en milieu marin (bleu) ou d'eau douce (vert) dans différentes régions du globe. Les fractions de taille ciblées sont toutes inférieures à 5  $\mu\text{m}$ . Les étoiles indiquent qu'au moins un représentant du groupe, de diamètre inférieur à 7  $\mu\text{m}$ , a été morphologiquement décrit (l'appartenance aux petits eucaryotes est donc visuellement vérifiée), en marin (bleu) ou eau douce (vert). D'après Fenchel, 1982; Stockner, 1988; Corpe and Jensen, 1992; Massana *et al.*, 2004; Seenivasan *et al.*, 2013. Ce bilan n'est évidemment pas exhaustif mais permet de mettre en évidence la grande diversité taxonomique des plus petits eucaryotes.

## Introduction

On notera cependant que les études de la diversité des petits eucaryotes par des méthodes moléculaires se basent sur des préfiltrations en séries, qui permettent en théorie de ne récolter que les cellules d'une fraction de taille voulue (souvent entre 0,2 et 2, 3 ou 5  $\mu\text{m}$ ). Or, ces filtrations ne sont certainement pas fiables à 100%, et une partie de la diversité détectée ainsi pourrait en fait correspondre à des organismes un peu plus grands que ceux ciblés (Massana 2011).

## 2) Abondance et importance écologique.

Les petits eucaryotes sont des composants abondants et très répandus des communautés planctoniques. Ils ont été détectés dans de très nombreux écosystèmes aquatiques, depuis les lacs d'eau douce (e.g. Richards and Bass, 2005) jusqu'aux grandes profondeurs océaniques (López-García *et al.* 2001) en passant par les écosystèmes marin côtiers (e.g. Romari and Vaultot, 2004), la glace (Piwoż *et al.* 2013) ou la surface des océans (e.g. Moon-van der Staay *et al.*, 2001). D'après des comptages en microscopie à épifluorescence et en cytométrie en flux, 1 ml d'eau de surface en milieu marin contiendrait entre  $10^2$  et  $10^4$  eucaryotes planctoniques de 1 à 3  $\mu\text{m}$  de diamètre (Massana 2011).

Les petits eucaryotes photosynthétiques notamment sont très abondants, et peuvent atteindre  $10^5$  cellules de moins de 3  $\mu\text{m}$  de diamètre par millilitre d'eau de surface riche en nutriments (Massana 2011). Leur rôle important dans la production primaire en milieu aquatique a été montrée très tôt : en 1994, Li montre que les petits eucaryotes ont un taux de fixation de  $^{14}\text{C}$  par cellule particulièrement élevé. Ils seraient responsables d'environ 40% de la fixation de  $\text{CO}_2$  en milieu marin (Jardillier *et al.* 2010).

Les petits eucaryotes hétérotrophes et mixotrophes quant à eux semblent être les principaux bactérivores dans les écosystèmes aquatiques (Zubkov & Tarran 2008; Massana 2011; Hartmann *et al.* 2012). De par leur taille intermédiaire, certains auteurs ont rapidement proposés que ces petits eucaryotes bactérivores fassent dans les réseaux trophiques le lien entre les bactéries phototrophes ou se nourrissant de matière organique dissoute, et les prédateurs de plus grande taille (Fenchel 1982; Pernthaler 2005).

Enfin, de nombreux petits eucaryotes vivent en interaction proche avec d'autres organismes, en tant que symbiontes (e.g. petits prymnesiophytes (Haptophytes) symbiontes de Rhizaria, Gast *et al.*, 2000) ou parasites (e.g. *Duboscquella*, Syndiniales, Harada *et al.*, 2007).

## II. Les eucaryotes unicellulaires en eau douce.

### A- Les milieux d'eau douce.

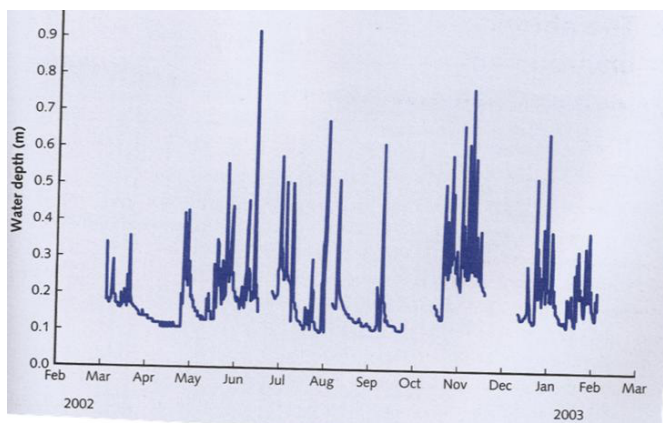
#### 1) Différents types d'écosystèmes d'eau douce.

Les écosystèmes aquatiques recouvrent la grande majorité de la surface terrestre, et, plus particulièrement, les écosystèmes continentaux couvrent plus de 3% des régions terrestres non océaniques (Downing *et al.* 2006). Bien que représentant un faible volume d'eau par rapport aux océans (Table 2), ces écosystèmes d'eau douce continentaux sont très divers et offrent donc une grande variété d'habitats. On les classe en différentes catégories selon leur dynamique et leur taille.

Réservoir	Volume ( $10^3 \text{ km}^3$ )	Volume (% eau totale)
Océans	1 338 000	96,5
Glaces	24 400	1,8
Sols	23 400	1,7
Lacs d'eau douce	105	0,008
Lacs salés	85	0,006
Atmosphère	13	0,001
Rivières	2	0,0001

*Table 2 : Répartition de l'eau dans les différents réservoirs. D'après Dobson and Frid, 2008.*

Les **ruisseaux, rivières et fleuves** tout d'abord, constituent les eaux courantes, ou milieux lotiques. Ils se caractérisent par un mouvement unidirectionnel de l'eau, déterminé par la gravité. Ces écosystèmes présentent des caractéristiques très variables le long de leur cours, avec des zones peu profondes à courant rapide et des zones plus profondes, larges et calmes. La vitesse du courant est un paramètre important influant sur la composition des communautés. Le débit et la profondeur des écosystèmes lotiques sont variables au cours du temps. Ils peuvent présenter des variations saisonnières (apport d'eau par fonte des glaces, fortes pluies), ou plus rapides, notamment dans le cas des ruisseaux, de plus petite taille (Figure 17). Le flux d'eau peut à certaines périodes s'arrêter totalement. A l'extrême, les régions les moins profondes des ruisseaux peuvent s'assécher périodiquement, aboutissant à la formation de flaques qui ne sont plus connectées entre elles.



*Figure 17 : Evolution de la profondeur d'eau dans un ruisseau au cours d'un suivi annuel. (Dobson & Frid 2008)*



## Introduction

Les masses d'eau douce, entourées de terre, qui ne circulent pas, ou dont l'écoulement est faible par rapport au volume d'eau contenu, sont quant à eux appelés systèmes lentiques. Ces écosystèmes sont traditionnellement classés en différentes catégories. Les **lacs** tout d'abord présentent une grande surface et une profondeur importante. Ils sont divisés en trois zones : la zone littorale, peu profonde, le long des berges, la zone limnétique, de surface, éclairée par la lumière du soleil, et la zone profonde située sous la zone limnétique et que la lumière n'atteint pas (Berg *et al.* 2009). Dans les régions tempérées, ils présentent une forte stratification thermique en été, quand les eaux chaudes de surface ne se mélangent pas avec les eaux plus froides en profondeur. Les seuls échanges entre surface et profondeur ont lieu en automne et au printemps, quand le changement de température des eaux de surface modifie leur densité qui devient supérieure (automne) ou inférieure (printemps) à celle des eaux profondes, et entraîne donc un brassage sur toute la colonne d'eau. Les échanges entre eaux de surface et sédiments sont donc saisonniers (Berg *et al.* 2009). Les lacs sont alimentés par la pluie, les eaux souterraines et souvent par des rivières qui s'y déversent. Les **étangs** sont moins profonds et de plus petite surface que les lacs, mais le continuum de tailles ne permet pas de proposer des valeurs seuils reconnues par tous les auteurs. Contrairement aux lacs, ils ne possèdent pas de zone profonde aphotique, ce qui permet aux embryophytes de se développer sur toute la surface des fonds ou presque (Oertli & Frossard 2013). Ils ne sont en général pas stratifiés, ou sur de très courtes périodes de temps, les eaux de surface et de profondeurs sont donc globalement homogènes. Les **mares** sont de petits étangs. Leur profondeur est généralement de quelques dizaines de centimètres, mais là encore les valeurs seuils de surface et de profondeur sont mal définies. Elles sont essentiellement alimentées par les eaux de pluies ou souterraines. Elles peuvent être isolées ou appartenir à des zones humides : marais, tourbières, prairies humides. De nombreuses mares sont artificielles, creusées dans un but décoratif ou comme réserve d'eau (e.g. abreuvement d'animaux) par exemple. D'autres se forment naturellement dans les creux du relief. Contrairement aux écosystèmes plus grands, elles ne sont en général pas permanentes. On notera que le vocabulaire anglais ne distingue pas mares et étangs, et les regroupe sous le terme « ponds ».

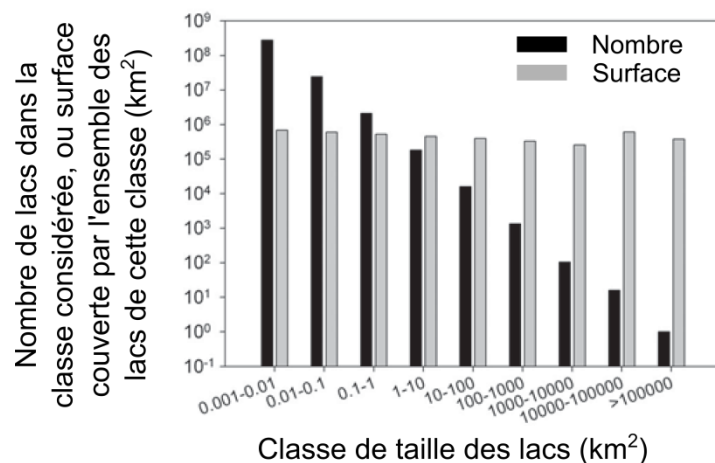
## 2) Les mares et ruisseaux, des écosystèmes particuliers

De par leur petite taille, les mares et ruisseaux constituent des écosystèmes aquatiques particuliers, très nombreux à la surface du globe. On estime que les petits étangs (moins de 1 000 m<sup>2</sup>) représentent plus de 91% de l'ensemble des écosystèmes continentaux lentiques (Downing *et al.* 2006). Il existerait 3 milliards de mares de 100 à 1 000 m<sup>2</sup> sur terre, inégalement réparties selon les régions, dont 1 million en France (Oertli & Frossard 2013). Le



nombre de mares étant bien plus élevé que celui des écosystèmes plus grands, la superficie continentale qu'elles recouvrent est similaire, elles offrent donc la même surface d'habitats que les grands lacs (Figure 18). Elles pourraient avoir une importance écologique considérable à l'échelle globale également. Il a été estimé par exemple que les mares artificielles de fermes piégeraient à elles seules plus de carbone que les océans, dans la matière organique sédimentée (Downing 2010).

**Figure 18 : Nombre et surface couverte par les lacs en fonction de leur taille.** (Downing 2010).

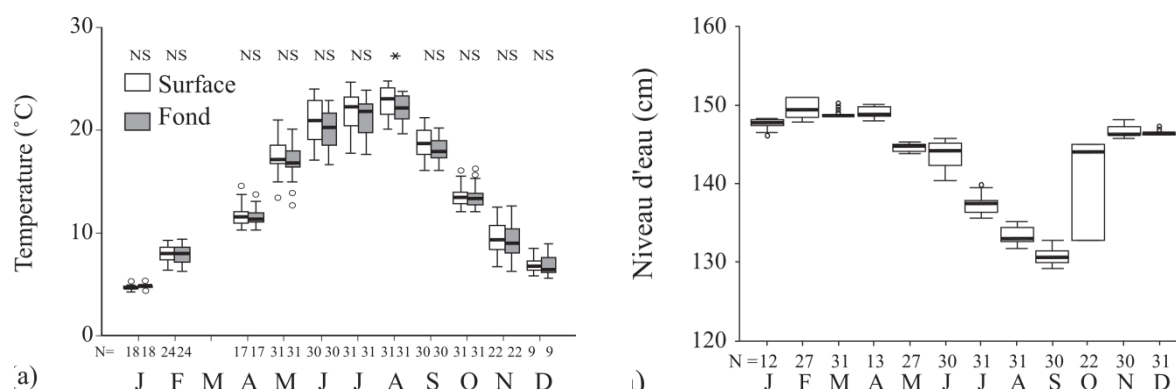


En plus d'être nombreux, les écosystèmes d'eau douce peu profonds sont également très divers. En effet, on en trouve en forêt, en plaine, en ville, en montagne, isolés ou en zones humides... Or, du fait de la faible inertie du petit volume d'eau qu'ils contiennent, et du rapport élevé surface de contact avec l'extérieur / volume, ces écosystèmes sont extrêmement influencés par leur environnement proche. La composition des sols et de la végétation alentour, la présence ou non d'activités agricoles dans le voisinage par exemple influencent fortement la composition de l'eau des mares et rus. D'autre part, les mares qui suivent leur évolution naturelle finissent par être comblées (on parle d'atterrissement). Entre la formation d'une mare et son atterrissement, elle connaît une succession écologique au cours de laquelle de nombreux paramètres évoluent, e.g. des différences de température et de teneur en oxygène peuvent apparaître entre la surface et le fond, les variations de niveau d'eau augmentent, et la végétation change (Angélibert *et al.* 2004). Ainsi, même dans une région géographique restreinte, on peut rencontrer des petits écosystèmes très variés, en fonction de leur environnement proche et de leur stade dans la succession écologique (Table 3).

Paramètre	Moyenne	+/- SD	Minimum	Maximum
pH	7,1	0,74	5	9,8
O <sub>2</sub> dissous (%)	56	42	0	158
Conductivité (µS/cm)	371	250	6,2	1367
NH <sub>4</sub> <sup>+</sup> (mg N/l)	0,21	0,47	0	3,58
NO <sub>2</sub> <sup>-</sup> (mg N/l)	0,01	0,017	0	0,112
NO <sub>3</sub> <sup>-</sup> (mg N/l)	0,74	1,14	0	7,7
PO <sub>4</sub> <sup>3-</sup> (mg P/l)	0,02	0,058	0	0,44
Transparence (cm Snellen)	42	19	4	60

*Table 3: Synthèse des résultats de la physico-chimie des eaux de 80 étangs de Suisse, montrant la grande diversité que l'on peut rencontrer au sein d'une même région géographique. SD: écart-type, leur grande taille montre les larges différences entre écosystèmes pour chaque paramètre (Oertli & Frossard 2013).*

L'une des caractéristiques essentielles des petits écosystèmes peu profonds, par comparaisons avec les lacs, plus grands et profonds, est la grande variabilité temporelle de leurs caractéristiques physico-chimiques. Ils sont en effet particulièrement sensibles aux variations de leur environnement proche : le vent entrainera un brassage des eaux sur toute la profondeur, une période de pluie apportera de l'eau et des nutriments par lessivage des berges. Les écarts de température de l'air ou les différences d'ensoleillement auront un impact direct sur toute la colonne d'eau, en raison de la faible profondeur. Dans le cours d'une journée d'abord, température, quantité de dioxygène dissous et pH peuvent fortement varier, en raison des variations de température de l'air et des différences du rapport activité photosynthétique / respiration cellulaire entre le jour et la nuit. Les mares et ruisseaux connaissent également de fortes fluctuations saisonnières. L'amplitude thermique annuelle de l'eau, suivant celle de l'air, est ainsi très importante (Figure 19). La concentration en nutriments varie également au cours de l'année, diminuant durant les saisons de forte croissance des organismes photosynthétiques. Mares et rus connaissent également de fortes variations saisonnières de leur profondeur (Figure 19), qui diminue lorsque la température augmente et la pluviosité diminue. Certains écosystèmes peuvent même s'assécher totalement de façon ponctuelle ou récurrente. Les petits écosystèmes peu profonds subissent également des variations interannuelles, liées à celles des facteurs climatiques ou dues à leur vieillissement, puisque les mares non entretenues par l'homme s'atterrissent progressivement jusqu'à disparaître. Ainsi, mares et ruisseaux proposent aux organismes qui y vivent des habitats très changeants.



**Figure 19 : Fluctuations saisonnières de la température (gauche) et du niveau d'eau (droite) dans une jeune mare des Causses, au cours d'un suivi annuel.** Les boxplots représentent la distribution des différentes mesures réalisées chaque mois (nombre de mesures  $N$  indiqué sous l'axe des abscisses). La différence entre les températures de surface et du fond est testée par un test de Mann-Whitney : NS: non significatif, \*:  $p < 0,05$ . Dans cette mare jeune, toute la colonne d'eau est homogène. (Angélibert et al. 2004).

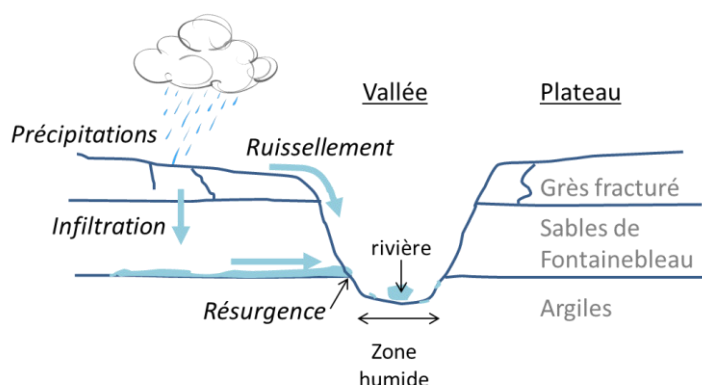
Outre leur grande variabilité et diversité, les petits écosystèmes peu profonds diffèrent des systèmes plus grands de par la grande surface de leur interface avec l'air, mais aussi avec les rives et les sédiments qui couvrent leur lit, par rapport au volume d'eau. Les petits écosystèmes peu profonds sont également peu ou pas stratifiés, et régulièrement brassés sur toute leur profondeur sous l'effet du vent par exemple. Ils échangent donc plus d'organismes et de composés chimiques avec l'atmosphère et le sol. Les macro-organismes qui y vivent sont également différents. Par exemple, les mares, non permanentes pour la plupart, n'hébergent pas de téléostéens, mais présentent une plus grande richesse et abondance d'embryophytes aquatiques (Downing 2010).

### 3) Les écosystèmes aquatiques dans le Parc Naturel Régional de la Haute Vallée de Chevreuse

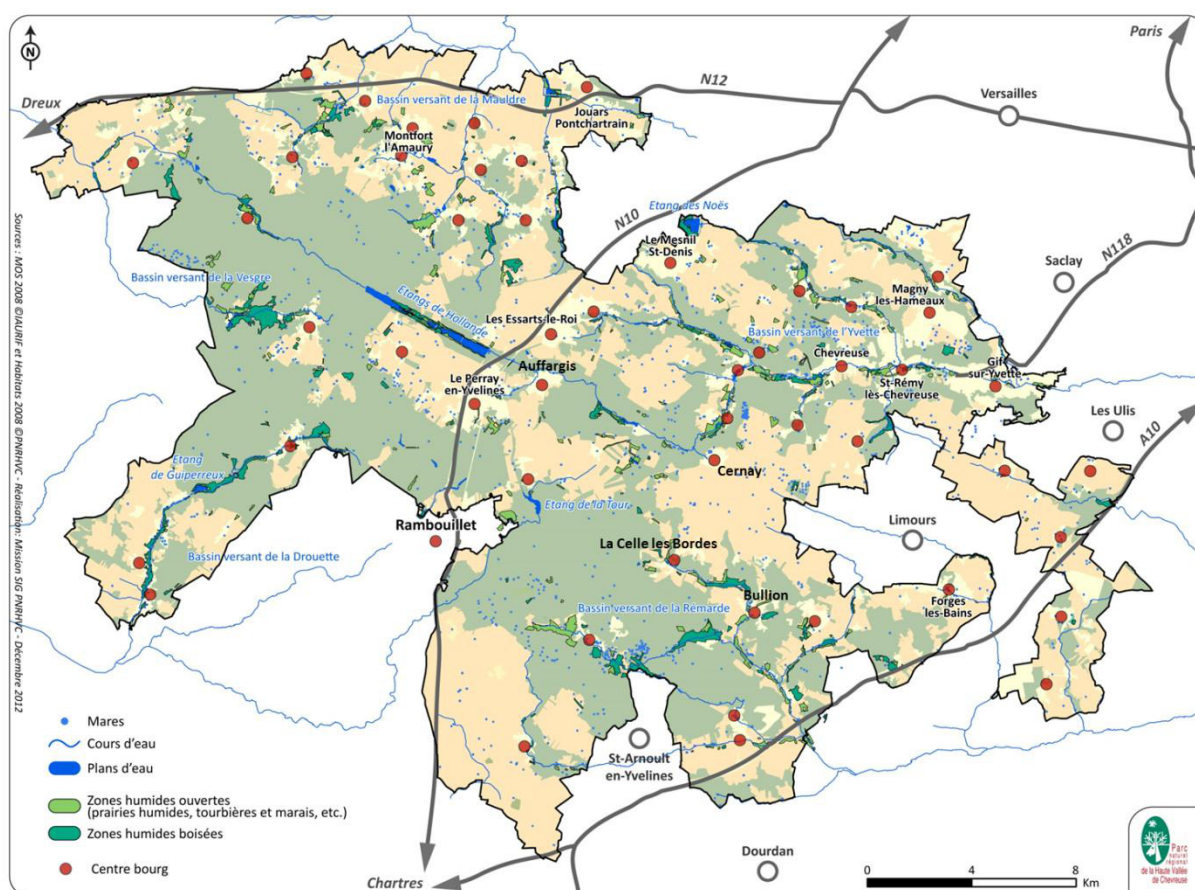
Le parc naturel régional de la haute vallée de Chevreuse (PNR HVC) est un territoire de 375 ha, situé au sud de Paris. Sa géologie est typique du bassin parisien : des plateaux de sables de Fontainebleau surmontés de bancs de grès reposent sur une couche argileuse. Ces plateaux sont entaillés par des vallées parfois encaissées formées par l'érosion due aux nombreux cours d'eau (<http://www.parc-naturel-chevreuse.fr/>).

Cette géologie particulière fait de ce territoire une région particulièrement riche en zones humides de toute sorte (Figure 20). En effet, les eaux de pluies ruissellent sur les plateaux ou s'infiltrent dans les sables qui forment une couche très perméable. Arrivées au niveau du socle argileux, très imperméable, elles ne peuvent s'infiltrer plus profondément et resurgissent à l'air libre au niveau des vallées, où se forment sources, mare, rivières...

**Figure 20 : La géologie du PNR explique la présence de nombreuses zones humides.**



Ainsi, le PNR estime que 5% de sa superficie est couverte par des zones humides. Il compterait notamment 330 km de rivières et quelques 1 800 mares (Figure 21). Ces dernières sont extrêmement variées, de par leur taille, leur environnement proche (forêts, villages, prairies, champs cultivés) ou l'origine de leur eau (sources, pluies, drainage artificiel).



**Figure 21: Plan du PNR HVC montrant la répartition des principaux écosystèmes aquatiques et zones humides.** (<http://www.parc-naturel-chevreuse.fr/nature-paysages/les-milieus-naturels.html>).

Ces milieux aquatiques et humides ont très tôt été modelés par l'action humaine, des zones humides ont été drainées pour être transformées en pâturages, des mares ont été creusées pour servir d'abreuvoir ou de réservoir d'eau pour l'arrosage. Notamment une chaîne

d'étangs et de rigoles, créée par Vauban au XVII<sup>ème</sup> siècle pour alimenter les bassins du château de Versailles, existe toujours aujourd'hui.

A l'heure actuelle, le PNR surveille ces écosystèmes de près. Notamment, de nombreux inventaires des macro-organismes animaux et végétaux sont réalisés. Différentes actions sont entreprises pour maintenir la diversité de ces milieux.

## **B- Diversité et aspects de l'écologie des eucaryotes unicellulaires en eau douce.**

### **1) Diversité des micro-organismes eucaryotes en eau douce.**

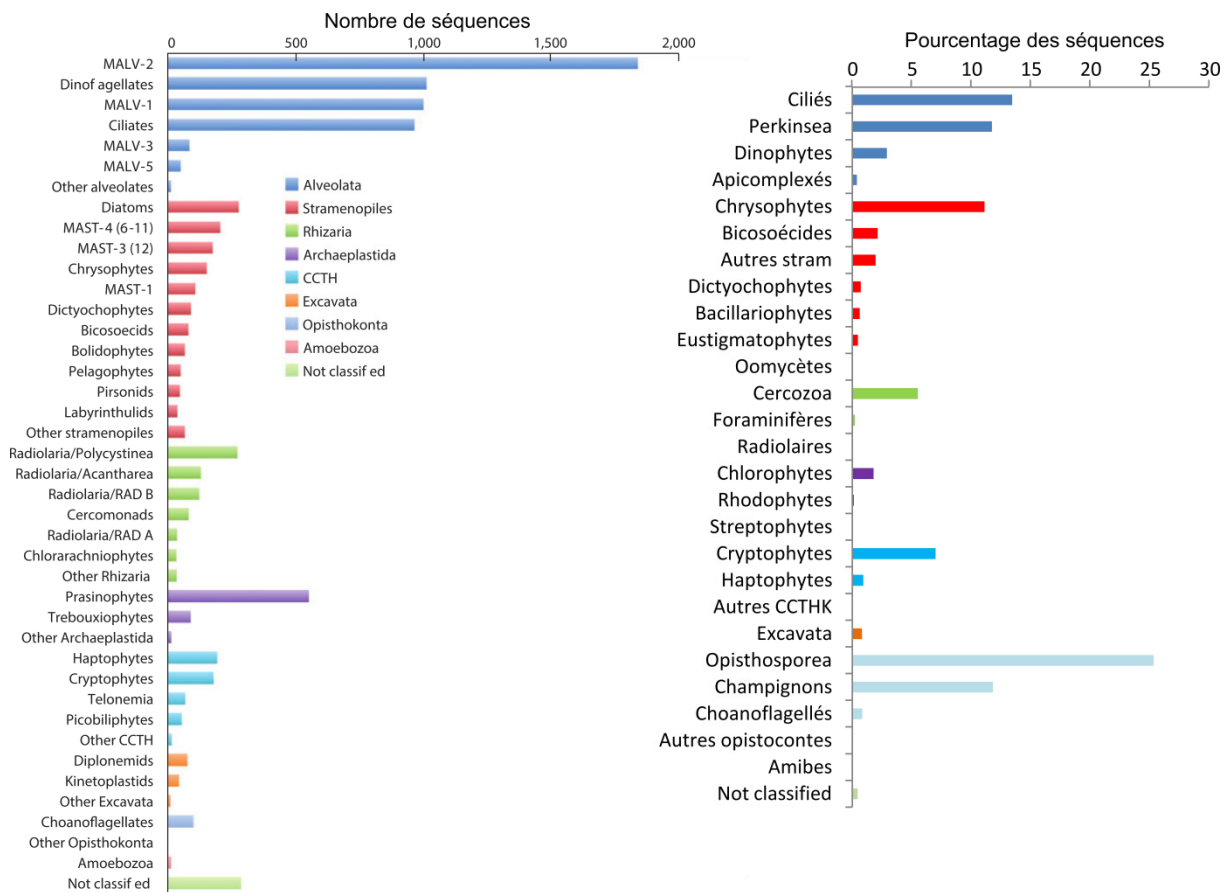
On rencontre dans les eaux douces des organismes appartenant à tous les supergroupes d'eucaryotes présentés plus haut, et au sein de ces grands groupes à un grand nombre de taxa de rang inférieur (Figure 22). La grande majorité des taxa présents en milieu marin sont également représentés en eau douce, mais généralement en des proportions différentes. On se concentrera ici sur les petits eucaryotes (<20 µm).

En eau douce comme en milieu marin, straménopiles et alvéolés sont parmi les supergroupes les plus abondants des communautés planctoniques. Cependant, alors que les séquences issues des écosystèmes marins sont essentiellement affiliées aux groupes de séquences environnementales MAST ou aux bacillariophytes (Diez *et al.* 2001; Massana 2011; Caron *et al.* 2012), les populations de straménopiles d'eau douce sont dominées, en terme de diversité et d'abondance, par les chrysophycées (Lefranc *et al.* 2005; Richards *et al.* 2005; Lepère *et al.* 2006; Lara *et al.* 2011; Triadó-Margarit & Casamayor 2012) et les bicosoécides (Richards *et al.* 2005; Lefèvre *et al.* 2008). Bien qu'en moindre abondance en général, d'autres taxa de straménopiles sont généralement rencontrés (e.g. oomycètes, bacillariophytes, labyrinthulides) (Šlapeta *et al.* 2005). Les alvéolés quant à eux sont en eau douce largement dominés par les ciliés (Šlapeta *et al.* 2005; Taib *et al.* 2013), qui représentent par exemple plus de 32% de l'ensemble des pyroséquences issues d'un échantillon du lac Léman en été (Mangot *et al.* 2012). Les dinoflagellés sont également souvent abondants (Richards *et al.* 2005; Lefèvre *et al.* 2008), et les perkinsés peuvent parfois atteindre de très fortes abondances relatives, ils représentent par exemple plus de la moitié des clones obtenus à partir d'un échantillon prélevé dans le lac Aydat en été (Lefranc *et al.* 2005). Cette grande abondance était passée inaperçue lors des observations microscopiques, et a été mise en évidence grâce aux méthodes moléculaires (Lefranc *et al.* 2005; Mangot *et al.* 2010). Les syndiniales, groupes extrêmement diversifiés et abondants en milieu marin n'ont en revanche jamais été détectés en eau douce (Guillou *et al.* 2008). Cryptophytes (Richards *et al.* 2005;



## Introduction

Šlapeta *et al.* 2005; Triadó-Margarit & Casamayor 2012; Taib *et al.* 2013) et champignons au sens strict, i.e. après exclusion des Rozellida-Cryptomycota (Lefranc *et al.* 2005; Lepère *et al.* 2008; Taib *et al.* 2013), sont également très souvent des membres majoritaires des communautés planctoniques d'eau douce, alors qu'ils sont plus rares en milieu marin (Massana 2011). On rencontre aussi régulièrement des membres du nouveau taxon Opisthosporidia (Karpov *et al.* 2014), i.e. séquences affiliées à LKM11 ou Rozellida-Cryptomycota, parfois en très forte abondance (Cf. affiliation taxonomique détaillée de Mangot *et al.*, 2012; Taib *et al.*, 2013). Les rhizariens sont presque exclusivement représentés par les cercozoaires, souvent plus divers qu'abondants (Taib *et al.* 2013). Contrairement à la surface des océans, les eaux douce hébergent relativement peu d'haptophytes, en dehors des épisodes d'efflorescences d'espèces généralement toxiques (Hansen *et al.* 1994).



**Figure 22: Abondances relatives de différents taxa eucaryotes en milieu marin (gauche) ou en eau douce (droite).** A gauche (Massana 2011), le graphe présente le nombre de séquences affiliées à chaque groupe dans des bibliothèques de clones construites après amplification d'ADNr 18S à l'aide d'amorces générales d'eucaryotes à partir d'échantillons marins de surface. 67% de ces séquences correspondent à des eucaryotes de moins de 3  $\mu\text{m}$ . A droite, l'histogramme présente l'abondance relative des séquences d'ADNr 18S affiliées à chaque groupe. Les valeurs indiquées correspondent à la moyenne des fréquences observées dans 6 lacs dont 2 échantillonnés à 2 dates différentes. Les séquences ont été obtenues par clonage – séquençage Sanger ou par pyroséquençage à partir d'échantillons de plancton correspondant à des cellules de 0,2 à 5  $\mu\text{m}$  de diamètre. (Lefranc *et al.* 2005; Mangot *et al.* 2012; Taib *et al.* 2013).

Ainsi, la plupart des taxa se rencontrent dans les deux types de milieux, bien qu'en des proportions souvent différentes. Cependant, au sein d'un taxon présent en eau douce et en milieu marin, les séquences d'ADNr 18S d'organismes d'eau douce forment généralement dans les phylogénies un nombre restreint de groupes, distincts des lignées marines (Logares *et al.* 2009). Cette observation a conduit à penser que les micro-organismes eucaryotes marins avaient conquis les écosystèmes d'eau douce depuis les océans lors d'un relativement faible nombre de transitions. Ce point sera discuté plus avant dans la discussion, au vu des résultats de ce travail de thèse.

D'un point de vue fonctionnel, on notera que les principales fonctions des réseaux trophiques ne sont pas réalisées par les mêmes groupes majoritaires en eau douce et en milieu marin. Alors que haptophytes, chlorophytes, bacillariophytes et parfois chrysophytes sont très abondants en milieu marin (Liu *et al.* 2009; Kirkham *et al.* 2011; Caron *et al.* 2012), les populations d'organismes photosynthétiques sont souvent dominées par les cryptophytes et chrysophytes en eau douce (Šlapeta *et al.* 2005). De même, alors que les syndiniales représentent un grand groupe de parasites en milieu marin, perkinsés et opisthosporidiens sont très abondants en eau douce (Bråte *et al.* 2010b).

## **2) Répartition géographique des eucaryotes unicellulaires en eau douce.**

Dès le début du XIX<sup>ème</sup> siècle, les scientifiques se sont intéressés à la biogéographie des macro-organismes, c'est-à-dire à la répartition de ces organismes à la surface du globe et aux déterminants de cette répartition. Notamment, Augustin de Candolle propose que la répartition des 'plantes' soit due à la fois aux conditions environnementales actuelles qui leur permettent ou non de se développer et à la géographie (isolement par la distance, existence par le passé de barrières géographiques ou au contraire de voie de migration entre deux lieux) (Martiny *et al.* 2006). Il est généralement admis que ces deux paramètres sont importants dans la répartition des macro-organismes. En revanche, la question de la biogéographie des micro-organismes et notamment des eucaryotes microbiens est fortement débattue depuis des années, et continue à faire couler de l'encre.

Selon la première hypothèse, les micro-organismes sont si petits et si abondants qu'il ne peut exister de barrière à leur dispersion dans le monde entier, via les courants océaniques ou atmosphériques, le vent, ou sur des animaux migrants par exemple (Finlay 2002; de Wit & Bouvier 2006). Si on n'observe pas les mêmes communautés partout, c'est que selon les conditions environnementales du lieu, des espèces différentes sont abondantes, et tous les autres organismes sont en dessous des limites de détection, actives ou sous des formes de résistance, voire en train de mourir (de Wit & Bouvier 2006). Cette théorie est en général



## Introduction

résumée par la célèbre phrase de Baas-Becking (1934) : « *Everything is everywhere, but the environment selects* ». Selon la vue opposée, il existe des barrières à la dispersion des micro-organismes, un isolement par la distance existe pour certains d'entre eux (Foissner 2006). Les deux parties se basent sur différents arguments:

→ *Cosmopolisme / endémisme* : Pour les partisans du rôle unique des conditions environnementales dans la distribution géographique des micro-organismes, aucune espèce microbienne n'est endémique d'une région géographique et elles peuvent donc être rencontrées en tout point du globe pourvu qu'on y trouve localement les conditions favorables à leur développement (de Wit & Bouvier 2006). Ils s'appuient notamment sur l'observation de morphoespèces identiques dans des régions distantes les unes des autres (Baas-Becking 1934; de Wit & Bouvier 2006), mais également sur la présence en des régions géographiquement lointaines de lignées quasiment identiques au niveau des marqueurs génétiques considérés (Šlapeta *et al.* 2006). A l'opposé, les partisans d'une distribution géographique des eucaryotes microbiens évoquent divers exemples d'endémismes (Foissner 2006). Il a ainsi été proposé que trois espèces de dinoflagellés soient endémiques du lac Baikal, ce serait aussi le cas de 8% des espèces de diatomées du lac Tanganyika (Foissner 2006). Le débat endémisme / cosmopolisme restera sans doute sans réponse. En effet, il semble parfaitement impossible de prouver que tout peut être observé partout comme de prouver que certains organismes sont endémiques d'une région. Il faudrait pour cela pouvoir décrire *en tout lieu* l'intégralité des communautés microbiennes, ce qui n'est pas réalisable. Même avec la grande profondeur de séquençage des nouvelles technologies permettant *a priori* de détecter des organismes en très faible abondance on ne peut détecter toute la diversité en un site ; ne pas détecter un organisme ne veut donc pas dire qu'il n'y n'est pas présent (peut-être à un autre endroit de l'écosystème considéré, ou à une autre période). Il semble donc difficile d'assurer qu'un groupe d'organismes est endémique d'une région.

→ *Fort/faible ratio diversité locale sur diversité globale*. Les partisans de l'absence de barrière géographiques à la dispersion, supposant que tout peut être trouvé partout, s'attendent à trouver à l'échelle du globe une diversité peu supérieure à celle que l'on peut rencontrer en une région donnée. Ils donnent plusieurs exemples dans ce sens. Notamment, Finlay observe 80% de l'ensemble des morphoespèces connues au sein du genre *Paraphysomonas* (Chrysophyceae-Synurophyceae, Stramenopiles) dans un seul étang d'1 ha (Green & Bohannan 2006). Cependant, la diversité globale n'est entièrement connue dans aucun taxon de micro-organismes eucaryotes, et ces observations sont généralement basées



sur des morphoespèces, qui pourraient cacher une plus grande diversité d'espèces cryptiques à l'échelle globale qu'au niveau local.

→ *Faible/forte relation taxa-surface*. Les partisans d'une différenciation des communautés par la distance supposent que plus des communautés seront géographiquement distantes, plus elles contiendront d'organismes différents. Ils s'attendent donc à rencontrer un plus grand nombre de taxa sur une plus grande surface, et une forte relation taxa-surface, définie par  $S = cA^z$  avec  $S$  le nombre de taxa,  $c$  une constante dépendant de l'environnement et du groupe considérés, et  $A$  la surface.  $z$  varie entre 0 (pas de relation entre le nombre de taxa et la surface échantillonnée) et 1 (relation très forte) (Smith *et al.* 2005). Certaines études ont mis en évidence des valeurs de  $z$  relativement élevées pour différents eucaryotes microbiens. Cependant, avec la surface augmentent à la fois la distance géographique entre les communautés et le nombre de niches écologiques potentiellement différentes. Aussi une forte relation taxa-surface peut être compatible avec l'existence d'un isolement par la distance comme d'une sélection par les conditions environnementales locales parmi des organismes mondialement dispersés.

→ *Corrélation distance géographique ou différences écologiques et similarité entre communautés*. Green et collègues (2004) ont montré que la similarité entre populations d'ascomycètes (Champignons, Opisthokonta) du sol en Australie diminuait quand la distance géographique entre ces populations augmentait. De même, Ragon et collègues (2012) ont montré que, contrairement aux bactéries, la distribution des eucaryotes dans des biofilms répartis dans différentes régions d'Europe, était essentiellement liée aux distances géographiques entre ces biofilms. Lepère et collaborateurs (2013) ont quant à eux testé la corrélation entre les différences de composition de communautés d'eucaryotes dans des lacs et la distance géographique, ainsi que la corrélation avec les différences environnementales entre ces lacs. Ils concluent que la distance géographique constitue la force principale influençant les différences de composition des communautés. Il semblerait donc qu'il puisse exister une structuration des communautés de micro-organismes eucaryotes par la distance, du moins dans des environnements fortement structurés comme les sols ou relativement isolés comme les lacs. Au contraire, des études ciblant spécifiquement les différentes lignées de *Spumella* (Chrysophyceae) ont montré que des génotypes d'ADNr 18S totalement identiques peuvent être trouvés dans des lacs de régions très éloignées mais au climat similaire (Nolte *et al.* 2010). En outre, des communautés de composition très différentes peuvent être observées dans des écosystèmes géographiquement proches (Lefranc *et al.* 2005).

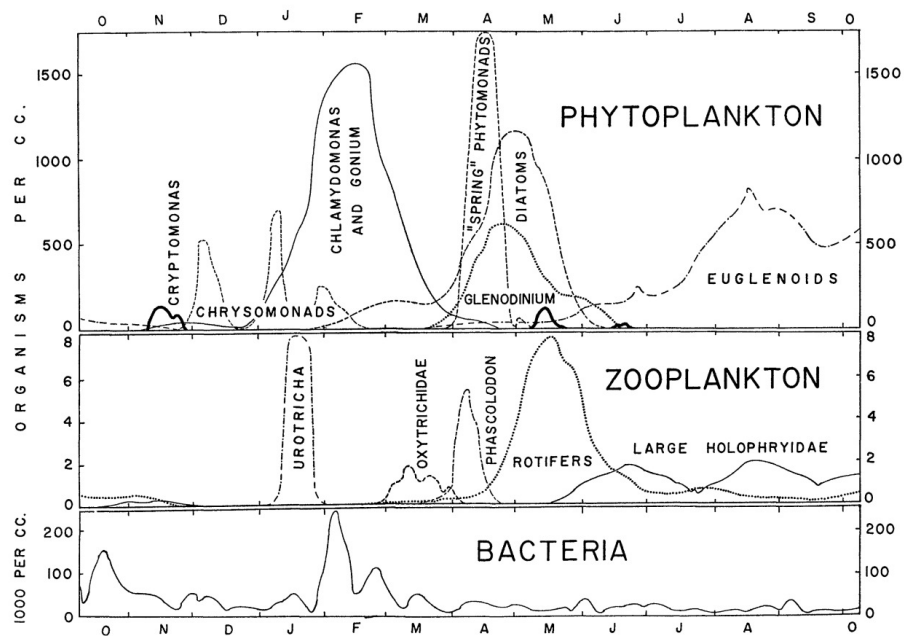
## *Introduction*

Ainsi, l'influence relative de la distance géographique et des conditions environnementales locales restent encore grandement à élucider. Les résultats souvent contradictoires obtenus par les méthodes présentées ci-dessus découlent peut être de la prise en compte d'échelles taxonomiques différentes, ou de la capacité de la méthode utilisée à détecter des taxa plus rares (chez de nombreux animaux et végétaux, les organismes les plus abondants sont aussi les plus rependus). L'effet relatif des deux forces dépend certainement des distances considérées, ainsi que des groupes taxonomiques (e.g. plus ou moins généralistes, présentant ou non des formes de résistance). Aussi, les travaux étudiant conjointement distances géographiques et caractéristiques environnementales locales devront être multipliés, à différentes échelles géographiques et taxonomiques, pour avoir une idée plus globale de l'impact des différents facteurs influençant les différences de composition des communautés de petits eucaryotes. Les écosystèmes continentaux semblent fournir un environnement adapté pour tester l'impact de ces facteurs. Ils représentent en effet un ensemble d'écosystèmes disjoints (il est nécessaire de passer par un milieu totalement différent pour aller de l'un à l'autre), comparables aux îles, dans lesquelles Darwin a mis en évidence la spéciation allopatrique (par isolement géographique) d'après l'observation des pinsons des Galapagos. Il serait plus facile d'y détecter d'éventuelles différences de composition des communautés en fonction de la distance que dans les océans où toutes les régions communiquent par le biais des circulations océaniques.

### **3) Dynamique des eucaryotes unicellulaires en eau douce.**

#### ***Premières observations.***

De la même manière que l'on observe des communautés de micro-organismes différentes en différents lieux, il a été noté depuis longtemps que des assemblages observés au même endroit en des moments différents ne sont pas identiques. Comme évoqué plus haut, Leeuwenhoeck remarque la dynamique temporelle des micro-organismes dès ses premières observations (Dobell 1932). Par la suite, les efflorescences périodiques d'organismes phytoplanctoniques (Nicholls *et al.* 1982), souvent remarquables même à l'œil nu, ont constitué des preuves claires de cette dynamique. Aussi, quelques suivis temporels de la diversité des eucaryotes microbiens basés sur des observations microscopiques ont été réalisés en milieu marin, mais également dans des lacs et mares d'eau douce. Par exemple, Bamforth suit la dynamique de différents groupes d'organismes planctoniques, en échantillonnant une petite mare un jour sur deux pendant un an (Bamforth 1958). Il observe ainsi des variations rapides et importantes de l'abondance de différents groupes d'eucaryotes (Figure 23).



**Figure 23 :** Succession des organismes dominants, au cours d'un suivi d'un an dans une petite mare artificielle (Pennsylvanie). Données recueillies par microscopie (Bamforth 1958).

L'arrivée des méthodes moléculaires a grandement facilité la réalisation de tels suivis, en permettant de déterminer plus rapidement la composition des communautés dans un grand nombre d'échantillons. Elles permettent en outre d'observer également la dynamique des organismes non identifiables morphologiquement. Notamment, grâce aux nouvelles technologies de séquençage telles que le pyroséquençage 454, on peut séquencer d'un même coup les amplicons d'ADNr 18S issues de plusieurs dizaines d'échantillons, pour un temps de manipulations relativement limité. D'autre part, la grande profondeur de séquençage de ces méthodes pourrait éviter d'observer des différences entre échantillons dues au séquençage d'un sous-ensemble aléatoire d'amplicons issus de chaque échantillon plutôt qu'à une réelle dynamique temporelle (Nolte *et al.* 2010). L'utilisation des méthodes moléculaires a ainsi permis de caractériser la dynamique des eucaryotes microbiens dans différents milieux, dont les écosystèmes continentaux d'eau douce.

### ***Les caractéristiques de cette dynamique.***

Les communautés d'eucaryotes microbiens des écosystèmes d'eau douce sont extrêmement variables au cours du temps ; des communautés prélevées mensuellement au cours d'une année peuvent ainsi présenter en moyenne plus de 92% de dissimilarité deux à deux (Jones *et al.* 2013). Cette dynamique est visible à différents niveaux. Elle concerne tout d'abord l'abondance totale du plancton eucaryote (Lepère *et al.* 2006; Mangot *et al.* 2012)

## Introduction

ainsi que la richesse et la diversité de leurs communautés. Jones et collègues (2013) montrent ainsi que le nombre d'OTU de séquences d'ADNr 18S, obtenues après pyroséquençage d'amplicons amplifiés à partir d'échantillons prélevés mensuellement dans un lac, peut être cinq fois plus élevé durant les mois d'été qu'en hiver. La dynamique concerne ensuite la composition des communautés d'eucaryotes microbiens. Il ne s'agit pas seulement de variations des proportions relatives des mêmes groupes ; certains groupes semblent également apparaître et disparaître au fil du temps (Nolte *et al.* 2010). Il semble clair que les phylotypes relativement abondants connaissent une grande dynamique temporelle (Lepère *et al.* 2006; Mangot *et al.* 2012). En revanche, la dynamique des organismes les plus rares, qui constituent la plus grande part de la diversité, est plus discutée. D'après Mangot et collègues, les taxa les plus rares varient peu entre les échantillons prélevés tous les 2 à 3 jours (Mangot *et al.* 2012), alors que pour Nolte et collaborateurs ces mêmes taxa sont les plus dynamiques au cours de leur suivi bimensuel (Nolte *et al.* 2010). Ces deux observations ne sont toutefois pas incompatibles puisque réalisées à différentes échelles de temps. Nolte *et al.* (2010) notent notamment que les phylotypes observés dans un échantillon unique sont toujours très peu abondants, et en déduisent que ces phylotypes très peu abondants ont une saisonnalité très marquée. Cependant, il est aussi possible que cette observation dans un seul échantillon soit due à la détection aléatoire d'organismes dont l'abondance fluctue peu mais reste toujours à la limite de détection, et pas du tout à de réelles variations d'abondance.

Cette forte dynamique de l'abondance, la diversité et la composition des communautés d'eucaryotes microbiens est observée à toutes les échelles de temps. On observe ainsi des communautés différentes dans des échantillons prélevés à quelques jours (Bamforth 1958; Mangot *et al.* 2012), quelques semaines (Nolte *et al.* 2010) ou quelques mois d'intervalle (Jones *et al.* 2013). Plusieurs études relèvent une dynamique saisonnière de l'abondance de certains groupes taxonomiques en eau douce : les chrysophytes semblent plus abondants en début d'hiver (Bamforth, 1958; Hitchman and Jones, 2000; Lara et al., 2011) alors que les populations de bacillariophytes paraissent croître au printemps (Bamforth 1958; Hitchman & Jones 2000; Mangot *et al.* 2009). Cependant, la plupart de ces études se basent sur des suivis d'une année ou moins, alors qu'une véritable saisonnalité se définirait par sa répétitivité d'une année sur l'autre. Certains suivis ont montré l'existence de cette cyclicité en eau douce (Jones *et al.* 2013) comme en milieu marin (Kim *et al.* 2014) : les communautés prélevées à un an d'intervalle semblent plus similaires que celles séparées de quelques mois seulement (et donc de saisons différentes). En revanche, d'autres études ont montré que la dynamique observée une année pouvait être totalement différente de celle observée l'année suivante (Lepère *et al.*

2006). La saisonnalité des communautés d'eucaryotes microbiens en eau douce reste donc une question en suspens.

### ***Facteurs environnementaux impliqués dans cette dynamique.***

De nombreux paramètres environnementaux sont susceptibles d'être impliqués dans la dynamique des communautés d'eucaryotes microbiens.

Les caractéristiques physico-chimiques de l'eau tout d'abord semblent responsables d'une partie de cette dynamique. Notamment, la température pourrait être un facteur déterminant. Il a ainsi été montré que c'était le principal facteur déclenchant le fort développement de *Goniostomum semen* (Raphidophyceae, Straménopiles) en été dans diverses mares d'Europe (Rengefors *et al.* 2012). La température expliquerait également à la fois la dynamique saisonnière et la distribution géographique des lignées de *Spumella* (Chrysophyceae, Straménopiles) (Nolte *et al.* 2010). Certains groupes d'eucaryotes pourraient aussi être sensibles à la teneur en oxygène dissous de l'eau (e.g. Cryptophytes Šlapeta *et al.*, 2005). L'intensité de la luminosité ainsi que la durée du jour influencent également la croissance des populations d'eucaryotes photosynthétiques. Le pH de l'eau, ainsi que le vent et l'hydrodynamisme du milieu pourraient également jouer un rôle dans la dynamique des communautés d'eucaryotes microbiens, directement ou par le biais d'autres facteurs sur lesquels ils influent, par exemple en augmentant la concentration en nutriments des eaux de surface par brassage (Lepère *et al.* 2006). Cette concentration en nutriments, notamment en phosphate et en composés azotés, constitue en effet un facteur prépondérant de régulation du phytoplancton. Il a été proposé que les petits eucaryotes photosynthétiques, de par leur plus grand rapport surface d'échange / volume, soient plus compétitifs que les organismes de plus grande taille, et semblent en effet plus abondants dans les milieux oligotrophes (Lepère *et al.* 2006), alors que des cellules de plus gros diamètre peuvent prendre le dessus quand la concentration en nutriments augmente.

La compétition entre organismes n'est pas la seule interaction biotique susceptible d'être impliquée dans la dynamique temporelle des communautés d'eucaryotes unicellulaires. La disponibilité des proies peut varier avec le temps, et entraîner une variation de l'abondance des prédateurs. La dynamique de *Cafeteria* (Bicosoécide, Straménopiles), bactérivore, paraît ainsi corrélée avec celle des bactéries dans le lac Pavin au cours d'un suivi bisannuel (Lepère *et al.* 2006). Dans le sens inverse, les variations d'abondance des prédateurs peuvent contrôler au moins partiellement la dynamique de leurs proies. Bamforth observe ainsi que le fort déclin de l'abondance de *Synura uvella* (Chrysophyceae, Straménopiles) est concomitant de la



## Introduction

forte augmentation de l'abondance de ciliés présentant des inclusions de chrysophytes (Bamforth 1958). Les organismes du zooplancton animal notamment sont d'importants prédateurs des eucaryotes unicellulaires et peuvent avoir un fort impact sur leurs communautés (e.g. Zhao et al., 2011). La dynamique des hôtes est également un facteur essentiel de la dynamique des organismes symbiotiques ou parasites. La présence de *Dinobryon* (Dinoflagellé, Alvéolés) serait ainsi responsable de 99,5% de la variation d'abondance des Perkinsozoa clade 2 expliquée par l'ensemble des paramètres pris en compte par Mangot et collègues (2009). Inversement, la dynamique des parasites peut fortement influencer celle des hôtes, jusqu'à réguler les efflorescences (Chambouvet *et al.* 2008).

L'équilibre dynamique entre populations virales et populations d'hôtes pourrait également être impliqué dans la variation temporelle des communautés d'eucaryotes microbiens en eau douce, comme ça a pu être observé pour les eucaryotes en milieu marin. En effet, il est régulièrement observé que des coccolithovirus sont en grande partie responsables de la disparition d'une efflorescence d'*Emiliania huxleyi* (Haptophyte) (Frada *et al.* 2008).

Tous ces paramètres sont susceptibles d'agir conjointement. Différentes méthodes statistiques peuvent permettre d'estimer quels paramètres parmi ceux pris en compte influencent le plus les communautés considérées. Les analyses canoniques de correspondance (CCA) sont particulièrement utilisées, et permettent de visualiser d'éventuelles corrélations entre des unités taxonomiques (OTU, genres ou groupes de rang taxonomique plus élevé) et un ou plusieurs paramètres environnementaux. Toutefois, même si ces analyses ont clairement un but explicatif, il faut garder en mémoire qu'on ne peut observer *in situ* que des *corrélations* entre paramètres environnementaux et dynamique des organismes. Or, ces corrélations peuvent témoigner de réelles relations directes de cause à effet, comme elles peuvent être fortuites ou dues à des relations indirectes. Une augmentation de la durée du jour pourrait par exemple être corrélée avec une forte croissance d'organismes purement hétérotrophes, uniquement parce que les jours plus longs permettraient le développement d'organismes phytoplanctoniques, et c'est ce développement qui serait la cause directe de la croissance des hétérotrophes. Les analyses réalisées *in situ* permettent donc d'identifier des responsables *potentiels* de la dynamique des communautés, mais ces résultats doivent être pris avec précaution, et mis en parallèle de ce que l'on connaît de la biologie des organismes considérés, ou testés *in vitro* en simplifiant le système et contrôlant tous les paramètres, avant de tirer toute conclusion définitive.



***Un cas particulier : la dynamique des communautés d'eucaryotes microbiens au cours d'un événement perturbateur ponctuel.***

Outre les variations quotidiennes ou saisonnières des conditions environnementales, certains événements surviennent ponctuellement au cours desquels l'environnement va être modifié fortement et brutalement. Ces événements perturbateurs peuvent être de toutes sortes. Il peut s'agir de l'efflorescence d'une espèce toxique (e.g. *Prymnesium parvum*, Haptophyte Jones *et al.*, 2013), d'un épisode climatique inhabituel ou particulièrement prononcé (e.g. sécheresse, fortes pluies), d'une pollution d'origine anthropique soudaine (e.g. épandage de pesticide à côté de l'écosystème aquatique juste avant de fortes pluies, déversement d'un composé toxique directement dans l'eau). Face à ces événements perturbateurs, les communautés peuvent être résistantes, c'est-à-dire ne pas être affectées par la perturbation, ou au contraire être sensibles et changer en réponse à la perturbation. Dans le second cas, elles peuvent être affectées durablement par la perturbation, ou revenir à l'état précédent l'événement perturbateur après que celui-ci se soit achevé. La résilience définit la vitesse de ce rétablissement, c'est l'inverse du temps de retour à l'état initial (Begon *et al.* 2009; Shade *et al.* 2012a). La plupart des quelques travaux ayant étudié la réponse des communautés d'eucaryotes microbiens à des événements perturbateurs en eau douce ont montré qu'elles étaient sensibles à ces événements (Shade *et al.* 2012a). Cependant, un suivi à plus long terme de leur dynamique, après que la perturbation ait disparu, semble montrer que ces communautés sont résilientes. Ainsi, au bout d'un certain temps, des communautés ayant subi le développement d'une efflorescence toxique (Jones *et al.* 2013) ou un important brassage du lac en plein milieu de sa période de plus grande stratification (Shade *et al.* 2012b) paraissent revenir dans un état proche de celui précédant la perturbation.

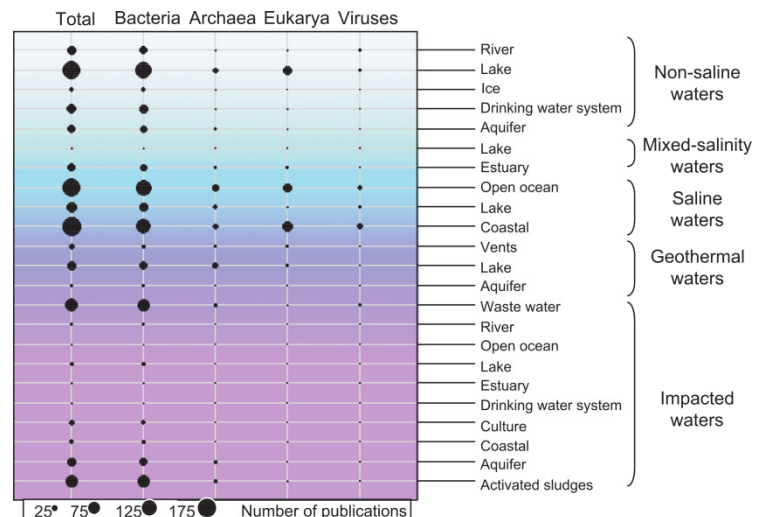
### **C- Les mares, des écosystèmes peu étudiés.**

Les études de diversité des micro-organismes eucaryotes aquatiques par des méthodes moléculaires ont commencé en milieu marin (Rappé *et al.* 1998; Diez *et al.* 2001; Moon-van der Staay *et al.* 2001; López-García *et al.* 2001). Par la suite, les premières explorations en milieu d'eau douce ont d'abord ciblé les environnements extrêmes. Ainsi, la première étude par des méthodes moléculaires réalisée en eau douce (2002) a ciblé un lac hyperacide (Amaral Zettler *et al.* 2002). Ce n'est que plusieurs années après les premiers travaux en milieu marin que des écosystèmes d'eau douce plus communs ont commencé à être explorés (Lefranc *et al.* 2005; Richards *et al.* 2005; Šlapeta *et al.* 2005). Aujourd'hui encore, tous les océans et mers du globe restent intensément étudiés, dans tous types d'environnements (e.g. profonds ou de

## Introduction

surface, côtiers ou de pleine mer, au niveau de sources hydrothermales) alors que le nombre de travaux réalisés en eau douce reste bien moindre. Pourtant, la diversité microbienne en eau douce pourrait être très large. Différentes études incluant à la fois des environnements marins et d'eau douce ont montré que la diversité des bactéries et archées était plus grande dans les écosystèmes d'eau douce continentaux, sans doute à cause de la moindre connectivité et de la plus grande hétérogénéité de ces systèmes (Zinger *et al.* 2011).

**Figure 24:** Nombre d'articles étudiant la diversité microbienne en milieu aquatique d'après le séquençage de gènes marqueurs, publiés entre 1990 et 2010, selon le domaine étudié et l'écosystème exploré. (Zinger *et al.* 2011).



D'autre part, l'essentiel des travaux conduits en eau douce l'a été dans de grands et profonds lacs, tels que le lac Georges ( $>100 \text{ km}^2$ , profondeur maximale 76 m, USA ; Richards *et al.* 2005), le lac Pavin ( $0,44 \text{ km}^2$ , profondeur maximale 95 m, France ; e.g. Lefranc *et al.* 2005; Taib *et al.* 2013), ou le lac du Bourget ( $42 \text{ km}^2$ , profondeur maximale 145 m France ; e.g. Lepère *et al.* 2008). Au contraire, les autres systèmes d'eau douce continentaux, étangs, rivières, eaux du sol, ont été laissés de côté (Figure 24, Zinger *et al.* 2011). Notamment, les petits écosystèmes peu profonds sont largement moins étudiés que les grands lacs.

Pourtant, tout porte à croire que la diversité microbienne en général et celle des eucaryotes en particulier pourraient être extrêmement importante dans les petits écosystèmes d'eau douce peu profonds, comme c'est le cas pour différents groupes de zooplancton animal et de macrophytes dont le nombre d'espèces par  $\text{km}^2$  diminue avec la superficie du lac (Downing 2010). Le grand nombre de ces écosystèmes et les grandes différences de l'un à l'autre offrent potentiellement une immense variété d'habitats différents. De plus, leur grande variabilité temporelle, et notamment les périodes d'assèchement total, pourrait sélectionner des organismes particulièrement résistants aux changements de leur environnement, que l'on ne rencontrerait pas nécessairement dans des écosystèmes plus stables. Les quelques études réalisées dans des mares confirment ce pronostic. Plusieurs études de diversité basées sur le séquençage d'ADNr 18S montrent l'existence d'une large diversité de micro-organismes eucaryotes, dans l'ensemble des supergroupes (Šlapeta *et al.* 2005; Lara *et al.* 2011). En

particulier, des organismes affiliés à des lignées basales au sein de différents taxons ont pu être détectés dans des mares. C'est le cas par exemple des Rozellida-Cryptomycota (Lara *et al.* 2010; Jones *et al.* 2011), branchant à la base des champignons et récemment rassemblés avec les aphélidés et microsporidies dans le groupe des Opisthosporidia, ou de la lignée HAP-1 qui semble être à la base des haptophytes (Šlapeta *et al.* 2005; Shalchian-Tabrizi *et al.* 2011). Cette dernière notamment n'a pour l'instant pu être détectée qu'en eau douce, et ne contient aucun organisme décrit.

Il se pourrait donc que les petits écosystèmes d'eau douce peu profonds, largement sous-explorés, hébergent une grande diversité de petits eucaryotes, dont toute une partie pourrait n'avoir encore jamais été observée.

### **III. Objectifs.**

#### **Améliorer notre connaissance de la diversité des petits eucaryotes.**

L'amélioration des phylogénies, de nos connaissances des relations de parenté entre groupes taxonomiques, passe entre autres par l'augmentation de l'échantillonnage taxonomique, donc par une meilleure connaissance de la diversité des organismes considérés. Le premier objectif de ce travail consistait en l'exploration d'une part sous étudiée de la diversité eucaryote, par deux approches un peu différentes :

Tout d'abord, nous nous sommes intéressés à la diversité intra-groupe au sein d'un taxon bien connu en milieu marin, mais dont la diversité a été très peu étudiée en eau douce : les haptophytes (chapitre 1). Afin de détecter la plus grande partie possible de cette diversité, nous avons ciblé le groupe d'intérêt en amplifiant un fragment de leur ADNr 18S à l'aide d'amorces *spécifiques* du groupe, avant clonage et séquençage Sanger des amplicons.

Ensuite, nous avons voulu explorer la diversité de l'ensemble des petits eucaryotes dans des écosystèmes où elle l'avait peu été jusque-là, c'est-à-dire des petits écosystèmes d'eau douce peu profonds. Nous avons pour cela séquencé des fragments d'ADNr 18S de l'ensemble des eucaryotes, par pyroséquençage d'amplicons obtenus à l'aide d'amorces générales d'eucaryotes, à partir d'échantillons de plancton issus de quatre mares et un ru (chapitres 2 à 4). Les communautés d'eucaryotes étant très variables, la diversité observable à un instant *t* ne représente qu'une fraction de celle hébergée par l'écosystème considéré. Afin d'avoir une vue plus complète de la diversité des eucaryotes dans les écosystèmes d'eau douce peu profonds, nous l'avons explorée au cours d'un suivi mensuel sur deux ans (chapitre 3 et 4).

## **Décrire la dynamique temporelle des communautés de micro-organismes eucaryotes dans les écosystèmes d'eau douce peu profonds.**

Les petits écosystèmes d'eau douce peu profonds étant peu étudiés, notamment en comparaison des grands lacs, l'écologie des communautés de micro-organismes eucaryotes qui y vivent reste à décrire. En particulier, la dynamique temporelle de ces communautés est peu connue, décrite seulement par quelques études basées sur l'observation microscopique des taxa les plus abondants. Pourtant, les caractéristiques physico-chimiques de ce type de systèmes étant très fluctuantes au fil du temps, on s'attend à une dynamique importante des communautés qui y vivent. Nous nous sommes donc attachés à décrire la dynamique temporelle des micro-organismes eucaryotes dans les écosystèmes d'eau douce peu profonds. Notamment, nous nous sommes demandé si cette dynamique présentait une saisonnalité (chapitre 1 et 3). Nous avons également eu l'occasion d'observer la dynamique des communautés lors d'un événement perturbateur majeur et avons donc voulu décrire la réaction des communautés planctoniques face à l'assèchement total de l'écosystème (chapitre 4).

## **Identifier des facteurs potentiellement impliqués dans la dynamique spatio-temporelle des communautés d'eucaryotes microbiens dans les mares et ru.**

Nous avons enfin voulu identifier des déterminants potentiels de la distribution spatiale et de la dynamique temporelle des eucaryotes microbiens dans les mares et ru.

Tout d'abord, les mares constituant des écosystèmes disjoints, aux caractéristiques environnementales variées, elles représentent un contexte de choix pour décrire la répartition géographique des micro-organismes eucaryotes, et tenter de déterminer le rôle relatif de la distance et des conditions environnementales dans la mise en place de cette répartition (chapitres 2 et 3).

Enfin, nous avons cherché à identifier d'éventuels paramètres environnementaux (caractéristiques physico-chimiques de l'eau et données météorologiques) qui pourraient être impliqués dans la répartition géographique et la dynamique temporelle des organismes étudiés (chapitres 2 et 3).

En particulier, la salinité est considérée comme un paramètre crucial dans la répartition des micro-organismes dans les écosystèmes aquatiques, qui limiterait fortement le nombre de transitions du milieu marin vers l'eau douce. En échantillonnant un large panel d'écosystèmes d'eau douce variés, et en comparant la diversité qu'on y rencontrait avec celle observée en milieu marin, nous nous sommes demandé si les différences de salinité constituent une si forte barrière à la dispersion, si ces transitions sont vraiment aussi rares que ce qui est généralement supposé (chapitre 1 et 2).

# Matériel et Méthodes

---



*"Good things come in small packages"*  
Proverbe anglais

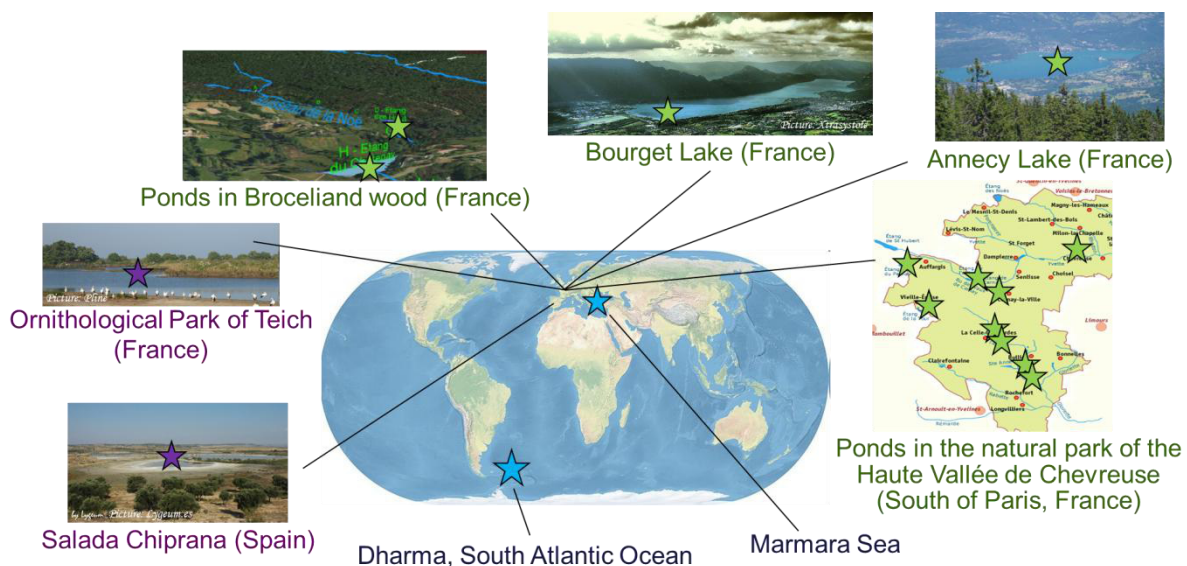




# Matériel et Méthodes

## I. Echantillonnage

Le premier volet de ce travail a consisté à comparer la diversité des haptophytes, à l'aide d'une nouvelle approche, dans des milieux marins, continentaux salés et d'eau douce. Dans cette optique, nous avons exploité les échantillons issus de 19 écosystèmes (prélevés pour cette étude ou disponibles au laboratoire, Figure 1), dont deux colonies d'eau marines géographiquement distantes (Atlantique Sud et mer de Marmara), deux milieux continentaux salés (La Salada de Chiprana hypersaline en Espagne (Jonkers *et al.* 2003) et une mare saumâtre du parc du Teich en Gironde), et 15 écosystèmes d'eau douce (lacs, mares et rus). Les différents échantillons prélevés dans ces systèmes et les modalités d'échantillonnage sont détaillés dans le chapitre 1 (Table S1).



**Figure 1: Localisation géographique des écosystèmes analysés.** Les systèmes d'eau douce, de milieux marins et continentaux salés sont indiqués par des étoiles vertes, bleues et violettes respectivement.

Par la suite, 5 de ces écosystèmes (Figure 2 et 3) ont été sélectionnés pour le suivi temporel de la diversité générale des petits eucaryotes. Ils ont été choisis d'après des informations données par le Parc Naturel Régional (PNR) de la Haute Vallée de Chevreuse, pour leur accessibilité mais surtout pour leurs caractéristiques contrastées. Nous avons ainsi étudié des représentants de types différents de milieu d'eau douce peu profonds du bassin parisien : une mare de forêt, une mare en contexte urbain, un étang, un ruisseau et une mare de zone humide boisée.



***Mare Gabard***

Coordonnées GPS 48°39'15.83"N - 1°55'20.26"E

Mare de forêt.

Prélèvement en surface à l'aide d'une perche et d'un bidon en plastique de 500 ml.



***Mare Saint Robert***

Coordonnées GPS 48°39'54.82"N - 1°56'45.28"E

Mare en contexte urbain (au sein d'un village).

Prélèvement en surface à l'aide d'une perche et d'un bidon en plastique de 500 ml.

***Etang des Vallées***

Coordonnées GPS 48°41'23,0"N - 001°54'59,2"E

Etang peu profond.

Prélèvement en surface au milieu de l'étang, depuis une barque, directement dans un bidon en plastique de 10 l.





***Ru Sainte Anne***

Coordonnées GPS 48°36'45.91"N - 1°58'16.61"E

Ruisseau à débit extrêmement variable.

Prélèvement en surface à l'aide d'un bidon en plastique de 500 ml.

***Mare de la Claye***

Coordonnées GPS 48°36'31.72"N - 1°56'17.33"E

Mare sur substrat tourbeux, appartenant à un complexe de mares dans une zone boisée.

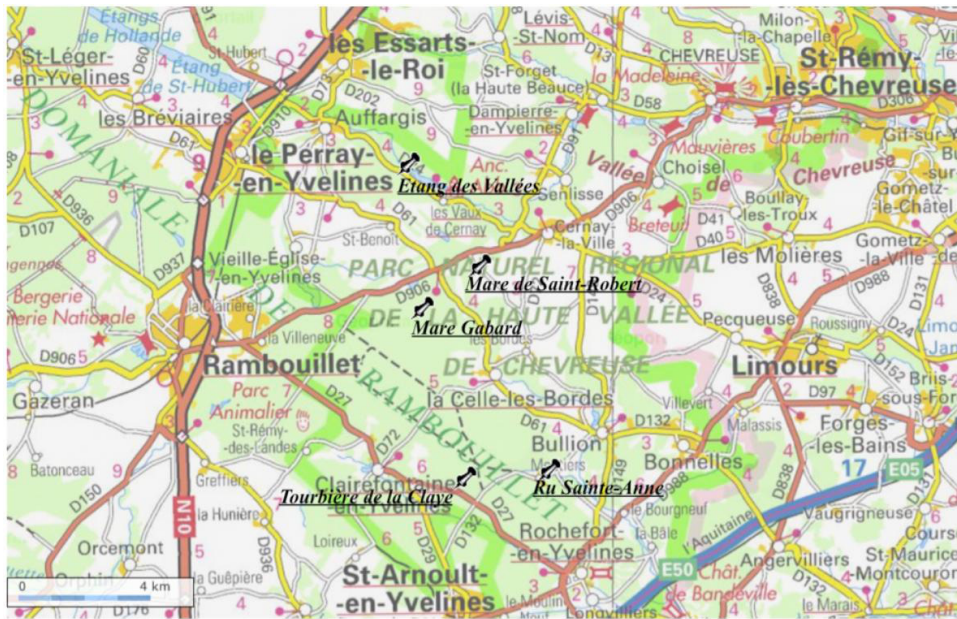
Prélèvement en surface à l'aide d'une perche et d'un bidon en plastique de 500 ml.



***Figure 2 : Les 5 écosystèmes d'eau douce peu profonds étudiés dans le cadre du suivi temporel de la diversité des eucaryotes.***

Les échantillonnages ont été réalisés chaque mois, sur trois jours, de mars 2011 à avril 2013. Environ 15 l d'eau par mois ont été prélevés dans chacun des écosystèmes. Lorsque les systèmes étaient à sec (La Claye de fin juillet à décembre 2011, puis en septembre 2012, Ru Sainte Anne en août et septembre 2012), du sédiment de surface a été prélevé au milieu de la zone habituellement submergée, dans des tubes Falcon stériles de 50 ml.





*Figure 3 : Localisation des cinq écosystèmes étudiés dans le cadre de l'étude de la diversité générale des petits eucaryotes. La frontière verte indique les limites du Parc Naturel Régional de la Haute Vallée de Chevreuse (carte IGN).*

## II. Préparation des échantillons

Les échantillons d'eau prélevés au laboratoire au cours de cette étude ont été préfiltrés sur 100 puis 30  $\mu\text{m}$  à l'aide de filtres de nylon (Millipore). Ils ont ensuite été fractionnés selon leur classe de taille par filtration différentielle sur des filtres Nucleopore (Whatman) de 5 puis 0,2  $\mu\text{m}$  de porosité. Ces filtrations ont été réalisées jusqu'à saturation des filtres (entre 50 et 500 ml environ sur les filtres de 0,2  $\mu\text{m}$ , jusqu'à 2 l sur les filtres de 5  $\mu\text{m}$ ). Les cellules étudiées par la suite ont été récoltées sur les filtres de 5  $\mu\text{m}$  de porosité (Etang des Vallées) et de 0,2  $\mu\text{m}$  de porosité (pour tous les écosystèmes d'eau douce). Au mois d'avril 2011, les filtrations ont été réalisées en réplicats pour les 5 écosystèmes décrits plus haut; les réplicats de filtres ont par la suite été traités indépendamment. Immédiatement après la filtration, les filtres sont stockés à  $-20^{\circ}\text{C}$  dans des tubes Eppendorf de 2 ml jusqu'à l'extraction d'ADN.

En cas d'absence d'eau (écosystèmes asséchés), les échantillons de sol ont été congelés à  $-20^{\circ}\text{C}$  dans des cryotubes de 5 ml jusqu'à l'extraction d'ADN. Un duplicat a également été réalisé en décembre 2011 à La Claye pour tester la reproductibilité de la méthode sur des sédiments.

Lors de l'étude de la diversité des haptophytes en eau douce, des échantillons de précédentes campagnes d'échantillonnage réalisées au sein de l'équipe ou dans d'autres laboratoires ont également été utilisés, afin de diversifier les milieux étudiés et leur origine

géographique. Certains n'ont pas été préparés suivant le protocole décrit ci-dessus. Les cellules planctoniques du lac d'Annecy et du Bourget ont été concentrées dans des Cell Traps (Mem-Teq) après pré-filtration des échantillons sur des filtres en polycarbonates de porosité 60 µm (fraction de taille théorique des cellules retenues : 0,22 – 60 µm de diamètre). Ces échantillons ont ensuite été plongés dans l'azote liquide avant d'être stockés à -80°C. Les cellules du parc du Teich analysées dans cette étude ont été obtenues par centrifugation d'un échantillon d'enrichissement en culture (K-medium (Keller *et al.* 1987) et Volvic).

### III. Extraction de l'ADN total

L'ADN total contenu dans les échantillons de sol ou sur les filtres a été extrait à l'aide du kit « PowerSoil DNA extraction » de MoBio. Pour extraire l'ADN des cellules recueillies sur filtre, la première étape du protocole a été adaptée : les filtres ont été découpés en petits fragments à l'aide d'un scalpel, dans une boîte de Pétri stérile, puis ajoutés dans les tubes à billes (lyse des cellules). Le protocole indiqué par le fabricant a ensuite été suivi à partir de l'étape 2. L'ADN a été élué dans 50 à 80 µl de tampon d'éluion du kit, puis aliquoté dans 2 tubes Eppendorf stockés à -20°C avant utilisation. L'ADN contenu dans les échantillons en CellTraps a été extrait d'après un protocole adapté de (Jardillier *et al.* 2010). Pour les culots de cellules en culture, l'ADN n'a pas été extrait : les PCR ont été réalisées directement sur ces culots rincés et resuspendus dans du Tris 10mM.

### IV. Amplification et séquençage des ADNr 18S

#### 1) Clonage et séquençage Sanger (diversité des haptophytes en eau douce)

L'ADNr 18S des haptophytes a été amplifié à l'aide d'amorces spécifiques de chacune des deux classes d'haptophytes : les prymnésiphytes et les pavlovophytes. Ces amorces (Table 1) ont été dessinées par David Moreira (Equipe Diversité, Ecologie et Evolution Microbiennes, ESE), puis sélectionnées parmi différents couples suite à des tests *in vitro* au laboratoire.

Les amplifications ont été réalisées dans un milieu réactionnel de 25 µl (1,5mM MgCl<sub>2</sub>, 1 µl DMSO (Sigma-Aldrich), 0,2 mM de chaque dNTP, 0,2 µM de chaque amorce, 0,5 à 4 µl d'ADN et 0,5 U de polymérase hotstart (Taq Platinum, Invitrogen). Les PCR ont été conduites en TouchDown (10 cycles de 15 s à 94°C, 30 s de 60 à 51°C et 2 min à 72°C, puis 25 cycles de 15 s à 94°C, 30 s à 55°C et 2 min à 72°C, pour finir avec une étape d'extension finale de 7 min à 72°C).

Des fragments des gènes de l'ARNr 18S des haptophytes ont pu être amplifiés par PCR directe à partir des échantillons marins. Pour les échantillons continentaux, il a été nécessaire de réaliser des PCR nichées. La première amplification a été réalisée à l'aide des amorces générales d'eucaryotes 82F et 1498R (López-García *et al.* 2003), et la seconde grâce aux amorces spécifiques de prymnésiphytes et pavlovophytes décrites plus haut, dans les mêmes conditions de PCR.

Les amplicons ainsi obtenus ont ensuite été clonés en utilisant le kit « TOPO TA cloning » (Invitrogen), en suivant le protocole indiqué par le fabricant. Pour chaque échantillon, 20 à 40 clones positifs ont été sélectionnés en général, et le fragment d'ADNr 18S qu'ils contenaient a été séquencé (Sanger) à partir de l'amorce forward utilisée pour l'amplification, par l'entreprise Beckman Coulter Genomics (Takeley, UK).

### 2) Pyroséquençage 454 (suivi de la diversité générale des eucaryotes)

Un fragment du gène codant l'ARNr 18S des eucaryotes, couvrant la région hypervariable V4, a été amplifié par PCR à l'aide d'amorces générales des eucaryotes biaisées contre les métazoaires, afin d'augmenter la proportion de séquences appartenant à des organismes unicellulaires dans notre jeu de données (Table 1). Nous avons utilisé 20 couples de ces amorces, chacun étiqueté grâce à la présence d'un Mid de 10 nucléotides ("Molecular identifier") différent en amont des séquences des amorces forward et reverse.

Les PCR ont été réalisées dans un milieu réactionnel de 25 µl (1,5 mM de MgCl<sub>2</sub>, 0,2 mM de chaque desoxynucléotide (PCR Nucleotide Mix, Promega), 0,3 µM de chaque amorce, 0,3 à 2,5 µl d'ADN et 0,5 U de polymérase hotstart (Taq Platinum, Invitrogen). Les amplicons ont été obtenus par 25 cycles d'amplification (30s à 94°C, 45 s à 58°C et 90 s à 72°C, pour finir sur 10 min à 72°C pour l'étape d'extension finale). Pour chaque échantillon, les produits de 5 à 11 PCR indépendantes ont été poolés.

Les amplicons poolés ont ensuite été purifiés à l'aide du kit « QIAquick PCR Purification » (Qiagen), en suivant le protocole indiqué par le fabricant.

Les amplicons purifiés de chaque échantillon ont été dosés au NanoDrop (ThermoScientific).

Les échantillons ont enfin été multiplexés : 20 échantillons amplifiés à l'aide d'amorces étiquetées par des Mids différents ont été poolés. Le volume de chaque échantillon a été déterminé d'après la concentration déterminée au NanoDrop de façon à ce que tous les échantillons soient présents à des concentrations théoriquement identiques dans le multiplex final.

Les multiplex ont ensuite été envoyés dans la carboglace à l'entreprise Beckman Coulter Genomics (Danvers, USA) où ils ont été séquencés par pyroséquençage 454 GS FLX Titanium (Roche).

Groupe ciblé	Utilisation	Amorce forward	Amorce reverse
<b>Eucaryotes</b>	1 <sup>ère</sup> amplification avant nichée.	EK-82F GAAACTGCGAATGGCTC	: 18S-1498R CACCTACGGAAACCTTGTTA
<b>Prymnesiophytes</b>	Pour clonage puis séquençage.	Pry421F AGCAGGCGCGTAAATTGCCCG	: Pry1572R TCAACGYRCGCTGATGACA
<b>Pavlovophytes</b>	Pour clonage puis séquençage.	Hap220F ACCGGTCTCCGGTTGCGTGC	: Pav1702R TAGATGATAAGGTTTGGGTG
<b>Eucaryotes (biaisés contre les métazoaires)</b>	Pour pyroséquençage 454.	EK-565F : GCAGTTAAAAAGCTCGTAGT	UNonMet : TTTAAGTTTCAGCCTTGCG

*Table 1 : Séquences des amorces utilisées dans les différentes études. Pour les amorces de pyroséquençage 454 (amorces générales d'eucaryote), les séquences des MIDs ne sont pas indiquées.*

## V. Nettoyage des séquences, construction des OTU et affiliation

### 1) Traitement des séquences issues du séquençage Sanger

L'affiliation au groupe des haptophytes des séquences fournies par l'entreprise de séquençage a tout d'abord été vérifiée par BLAST contre la base de données SILVA SSU104 (Pruesse *et al.* 2007).

Les séquences d'ADNr 18S d'haptophytes ont alors été vérifiées manuellement par observation des chromatogrammes à l'aide du logiciel CodonCode Aligner. Les extrémités des séquences correspondant à un chromatogramme de mauvaise qualité ont été éliminées, aboutissant à des séquences d'environ 700 pb.

La présence de séquences chimériques a été recherchée en utilisant le logiciel en ligne KeyDNATool (<http://KeyDNATools.com>) ainsi que manuellement en blastant indépendamment les moitiés 3' et 5' des séquences contre la base de données Silva SSU104.

Toutes les séquences considérées comme non chimériques ont alors été coupées à la même position dans CodonCode Aligner, puis une matrice de distance entre ces séquences a été construite avec ClustalX (Larkin *et al.* 2007). Cette matrice a permis de regrouper les séquences en OTU à 98% d'identité à l'aide du logiciel Mothur (Schloss *et al.* 2009). La valeur de 98% d'identité a été choisie d'après Caron and Countway (2009), et correspondrait au pourcentage de similarité moyen entre séquences d'ADNr 18S au sein d'une « espèce » chez les microorganismes eucaryotes. En construisant des arbres phylogénétiques rapides incluant toutes nos séquences (alignement avec Mafft, puis reconstruction phylogénétique en maximum de vraisemblance (ML) avec TreeFinder), nous nous sommes aperçus que quelques OTU construites par Mothur étaient polyphylétiques. En effet, Mothur se base exclusivement sur une matrice de distance et donc ne prend en compte que le *nombre* de différences entre deux séquences et non la nature de ces différences, ce qui peut dans certains cas aboutir à des OTU incohérentes avec les relations phylogénétiques entre les séquences. Aussi, quelques OTU ont été redécoupées manuellement. Par la suite, le terme « OTU » pour les séquences issues du séquençage Sanger correspondra donc à un groupe de séquences partageant au moins 98% d'identité et formant un groupe monophylétique.

Pour chacune de ces OTU, une séquence représentative a été choisie puis l'amplicon correspondant a été séquencé en sens inverse à partir de l'amorce reverse utilisée pour l'amplification. L'assemblage des séquences forward et reverse (CodonCode Aligner) a permis l'obtention de séquences complètes d'environ 1400 pb pour les pavlovophytes et 1070 pb pour les prymnésiphytes.

Une première affiliation des OTU aux différents taxons d'haptophytes a été réalisée en blastant les séquences complètes représentatives de chaque OTU contre la base de données Silva SSU104.

## 2) Traitement des séquences issues du pyroséquençage 454 FLX Titanium

L'entreprise de séquençage (Beckman Coulter Genomics) fournit les séquences obtenues lors du run de séquençage après démultiplexage mais sans aucun autre traitement ni tri. Or, environ 33% des séquences issues de pyroséquençage 454 GS-FLX Titanium contiennent des erreurs de séquençage sur les 101 premières paires de bases, et 90% sur leur longueur totale (Gilles *et al.* 2011). Un tri des séquences brutes est donc indispensable pour éliminer celles produites artefactuellement lors du séquençage.

Ce tri a été réalisé en collaboration avec Philippe Deschamps (Equipe Diversité et Evolution Microbienne, ESE), grâce à une pipeline de sa conception mettant en œuvre une



procédure testée par comparaison avec d'autres méthodes de traitement des données issues du pyroséquençage 454 (Bachy *et al.* 2013). Cette procédure consiste en plusieurs étapes. Toutes les séquences sont traitées ensemble indépendamment de l'échantillon dont elles proviennent. Tout d'abord, un prétraitement est appliqué qui consiste en une élimination des séquences contenant des erreurs dans les amorces ou des positions auxquelles la nature du nucléotide n'a pas pu être déterminée, ainsi que les séquences très courtes ( $< 450\text{pb}$ ). Dans certaines études, les séquences non complètes (absence de la séquence d'une amorce) sont éliminées des jeux de données ; elles ont ici été conservées afin de prendre en compte les séquences contenant des insertions, et donc trop longues pour être entièrement séquencées. Les erreurs de PCR et de pyroséquençage sont éliminées à l'aide d'AmpliconNoise (Quince *et al.* 2011) dans l'ensemble des séquences conservées à l'issue de ce prétraitement. Les séquences ainsi nettoyées sont regroupées en OTU à 98% d'identité par agrégation suivant la distance moyenne (=Average linkage clustering) après alignement des séquences deux à deux (AmpliconNoise). Les singletons (OTU ne contenant qu'une seule séquence) sont éliminés. Chaque OTU est représentée par sa séquence la plus abondante. On utilise alors ces séquences représentatives pour déterminer l'affiliation de chaque OTU en les blastant contre les bases de données de séquences d'ADNr 18S SILVA SSU111 (Pruesse *et al.* 2007) (chapitre 2) ou PR2 (Guillou *et al.* 2013) (chapitre 3 et 4). Les séquences constituant ces OTU sont attribuées à l'échantillon dont elles proviennent grâce aux MIDs étiquetant les amorces. Chaque OTU est donc caractérisée par une séquence représentative, une affiliation taxonomique, un score lié au pourcentage de couverture du premier hit par blast et le nombre de séquences de cette OTU dans chaque échantillon.

Les OTU étudiées dans le chapitre 2, détectées dans les échantillons prélevés en avril 2012, ont été construites à partir des reads issus du premier run de séquençage. Pour la suite du projet (suivi de la diversité pendant 24 mois et observation de l'impact des périodes de sécheresse sur les communautés planctoniques), les OTU ont été construites de nouveau à partir des reads issus de 3 runs de pyroséquençages différents.

Quelques blasts manuels contre la base de données Nucleotide Collection (nr/nt) sur NCBI ont révélé la présence d'une quantité non négligeable de chimères parmi les séquences représentatives des OTU. Pour éliminer les séquences potentiellement chimériques, nous avons mis au point le processus schématisé dans la figure 4. Les séquences ayant moins de 90% de couverture avec leur premier hit par blast (blast réalisé pour l'affiliation des OTU, Cf. ci-dessus) ont été éliminées de la base de données étudiée par la suite si elles correspondaient à des OTU représentées dans un seul échantillon. Si elles étaient représentées dans au moins

## *Matériel et méthodes*

deux échantillons, leur affiliation manuelle a été notée "Uncertain". En effet, on considère alors qu'elles ne sont pas suffisamment proches de leur premier hit pour que l'on puisse affirmer qu'elles appartiennent au même groupe taxonomique. Les séquences ayant au moins 90% de couverture avec leur premier hit par blast ont dans un premier temps été vérifiées à l'aide du logiciel KeyDNATool (<http://KeyDNATools.com>). Les bases de données utilisées par ce logiciel n'étant pas extrêmement riches en séquences eucaryotes, beaucoup de séquences étaient artefactuellement considérées comme chimériques. Par conséquent, les séquences considérées suspectes par KeyDNATool ont été vérifiées manuellement en blastant indépendamment les moitiés 3' et 5' ou les trois tiers des séquences contre la base de données PR2 grâce à un script écrit au laboratoire par David Moreira. Les séquences représentant des OTU présentes dans au moins deux échantillons ou dont les blasts des différentes parties étaient cohérents ont été conservées. Lorsque les résultats des blasts n'étaient pas cohérents, les séquences ont été soit directement éliminées de la base de données étudiée par la suite (les deux moitiés de séquences blastant avec des séquences appartenant à des groupes phylogénétiquement distants i.e. dans des supergroupes différents), soit vérifiées manuellement à l'aide d'un blast contre la base de données Nucleotide Collection (nr/nt) sur NCBI. Une grande partie des séquences issues du premier run a été vérifiée manuellement. Une comparaison rapide des séquences identifiées comme chimériques manuellement ou à l'aide du processus décrit ci-dessus nous a montré que la quasi-totalité des chimères semble détectée par le processus partiellement automatisé. En revanche, ce processus est certainement trop sévère et élimine des séquences réelles. Notamment, parmi les séquences ayant moins de 90% de couverture avec leur premier hit par blast éliminées, une partie correspond certainement à des chimères ou d'autres artefacts expérimentaux, mais certaines peuvent être des séquences divergentes sans séquences proches dans les bases de données. Nous avons cependant choisi la prudence en les mettant de côté. Les OTU correspondantes et toutes les informations associées restent toutefois disponibles, et pourront éventuellement être reprises dans un second temps pour une analyse plus poussée.

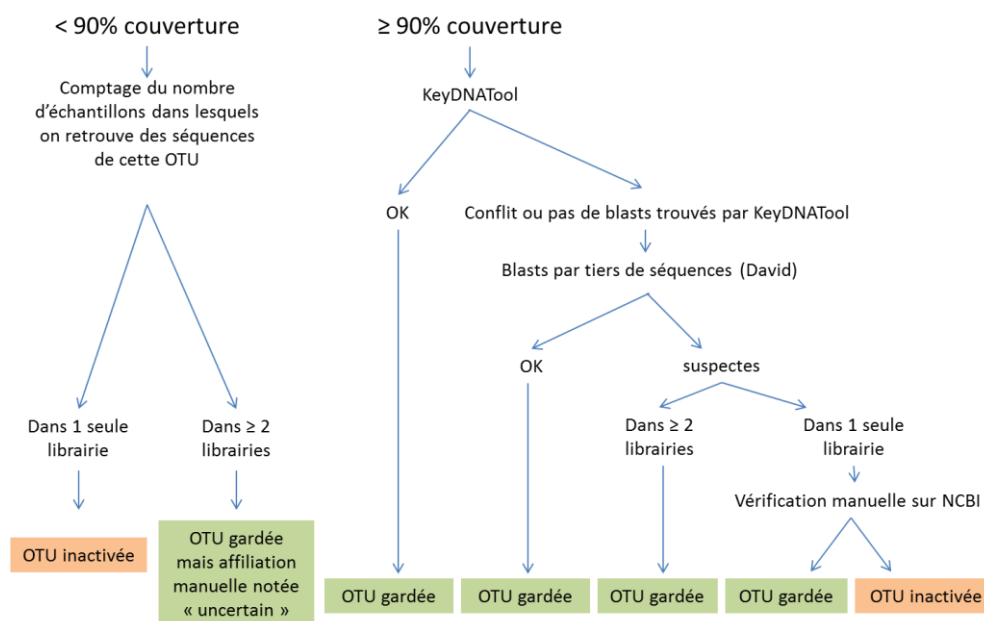


Figure 4 : Procédé semi automatisé de vérification des chimères.

Le jeu de données que nous avons obtenu contenait des séquences d'ADNr 18S de nucléomorphes de cryptophytes. Les OTU correspondant à ces séquences ont été éliminées de la base de données étudiée par la suite, afin de ne pas augmenter artificiellement l'abondance des cryptophytes.

## VI. Mesures des paramètres physico-chimiques

Ces paramètres ont été estimés dans les cinq écosystèmes étudiés pour le suivi de la diversité générale des eucaryotes, chaque mois, en parallèle du prélèvement des échantillons.

La température, le taux d'oxygène dissous, le pH et la conductivité ont été mesurés *in situ* à l'aide d'une sonde multiparamétrique WTW – Multi 350i.

La concentration de différents composés chimiques a été dosée au laboratoire, le jour du prélèvement, à partir d'un échantillon d'eau préfiltrée sur un filtre polycarbonate de 0,2 µm de porosité (Millipore ou Whatman) de façon à ne conserver que les substances dissoutes et les très fines particules. Ont été dosés durant la première année : l'azote dissous total, le nitrate, le nitrite, l'ammoniaque, le phosphore total, le phosphate, les carbonates, le carbone organique dissous, et la silice. La dureté de l'eau a été estimée également. Durant la première année, les concentrations en silice, ammoniaque et carbonate et la dureté ont très peu varié. On a alors cessé de les mesurer durant la seconde année d'échantillonnage. Toutes ces mesures ont été réalisées par colorimétrie à l'aide de tests Hach-Lange en sachet (Phosphate, silice, nitrite, ammoniaque) ou en cuve (Azote total, nitrate, phosphore total, carbonates,

## *Matériel et méthodes*

carbone organique dissous, dureté). Les mesures d'absorbance ont été réalisées avec le spectrophotomètre DR5000 (Hach-Lange).

Dans l'étude de la dynamique temporelle des communautés de petits eucaryotes (chapitre 3), des données météorologiques ont également été prises en compte. Elles ont été obtenues auprès de Météo France (<https://donneespubliques.meteofrance.fr/>). La pluviométrie a été estimée par la moyenne des mesures réalisées dans les trois stations météorologiques les plus proches: Choisel (48°41'00"N-2°00'00"E), le Perray (48°41'36"N-1°52'00"E) et Saint Arnoult (48°34'18"N-1°55'54"E). La pluviométrie associée à chaque échantillon correspond à la somme des hauteurs de pluie quotidiennes mesurées durant la semaine précédant l'échantillonnage (chapitre 3). L'évolution quotidienne des précipitations entre le dernier échantillonnage d'un écosystème asséché et le premier échantillonnage après la remise en eau a également été suivie (chapitre 4). Les moyennes des valeurs de rayonnement global quotidien et des températures quotidiennes sous abri au cours de la semaine précédant l'échantillonnage, ainsi que la moyenne des vitesses du vent à 2 m le jour de l'échantillonnage, ont été mesurées à la station météorologique de Trappes (48°46'24"N-2°00'30"E).

## **VII. Analyse des données issues du pyroséquençage 454**

Toutes les analyses statistiques ont été réalisées à l'aide du logiciel R (R Development Core Team 2013).

### **1) Comparaison des communautés : estimation du degré de ressemblance entre échantillons**

#### ***Calcul des indices de dissimilarité de Bray-Curtis***

Bien que les différents produits PCR aient été multiplexés en concentrations théoriquement équivalentes, le nombre de séquences obtenues pour chaque échantillon varie énormément (de 368 à 85 803 reads), même entre échantillons séquencés dans le même quart de run. Dans certains cas, le nombre de séquences obtenues était si faible que de nouvelles PCR ont été réalisées sur les quelques échantillons concernés afin de séquencer plus de matériel et obtenir des reads supplémentaires. Plusieurs hypothèses peuvent être avancées pour expliquer ces différences. Tout d'abord, les dosages au NanoDrop des amplicons purifiés sur lesquels je me suis basée pour calculer la quantité de chaque échantillon à pooler dans le multiplex peuvent être biaisés, par exemple lorsque des molécules autres que l'ADN restent dans la solution. Ensuite, le séquençage en lui-même semble aboutir à des nombres de séquences différents selon les échantillons, il a été proposé par exemple que tous les

amplicons ne soient pas aussi bien séquencés selon le Mid qui les flanque. Pour limiter le biais dû à ces différences de profondeur de séquençage lors de la comparaison des échantillons, les comptages bruts du nombre de séquences de chaque OTU dans chaque échantillon sont remplacés par la fréquence de chaque OTU dans l'échantillon considéré.

Le degré de différence entre deux communautés a été estimé grâce à l'indice de dissimilarité de Bray-Curtis (1957), calculé sur la fréquence des différentes OTU d'après la formule suivante :  $d = \frac{\sum_i |x_i - y_i|}{\sum_i (x_i + y_i)}$  avec  $x_i$  et  $y_i$  les fréquences de l'OTU<sub>i</sub> dans

les échantillons X et Y, respectivement. Il varie entre 0 (échantillons identiques) et 1 (échantillons n'ayant aucune OTU en commun). Les dissimilarités de Bray-Curtis ont été calculées avec la fonction 'vegdist' du paquet R 'Vegan' (Oksanen *et al.* 2013).

Cet indice est plus influencé par les OTU très abondantes dans au moins un des deux échantillons que par des OTU moins fréquentes. Ainsi, une OTU abondante partagée dans des proportions similaires par les deux échantillons diminuera beaucoup plus la dissimilarité entre les deux échantillons qu'une OTU partagée dans la même faible fréquence. En 1957, Bray and Curtis proposent une double standardisation des données qui permet d'équilibrer l'impact des OTU peu et très abondantes : la standardisation de Wisconsin. Elle consiste à standardiser dans un premier temps les fréquences de chaque OTU dans l'ensemble des échantillons en fonction de sa fréquence la plus élevée (i.e. pour chaque OTU, sa fréquence dans chaque échantillon est divisée par la fréquence de cette OTU dans l'échantillon où elle est la plus abondante). Puis les fréquences standardisées des différentes OTU au sein de chaque échantillon sont divisées par la somme des fréquences de toutes les OTU au sein de l'échantillon. Cette standardisation a été utilisée dans certains cas avant le calcul des indices de Bray-Curtis, notamment dans les représentations de Non-metric MultiDimensional Scaling (NMDS), pour comparer l'effet des OTU abondantes et moins abondantes.

### ***Représentations des indices de dissimilarité de Bray-Curtis***

Différentes représentations graphiques ont été utilisées pour représenter les dissimilarités de Bray-Curtis entre communautés. Les Non-metric MultiDimensional Scaling (NMDS), réalisées à l'aide du paquet R 'vegan' permettent de visualiser graphiquement quels ensembles d'échantillons se ressemblent le plus, et au contraire lesquels sont plus marginaux. Les NMDS permettent notamment de visualiser si les échantillons se regroupent en fonction

## Matériel et méthodes

d'un paramètre donné (e.g. entre répliqués d'un même échantillon, entre échantillons d'un même écosystème, d'une même saison). Elles ont été réalisées après transformation racine carrée et standardisation de Wisconsin des fréquences des OTU. L'importance de ces regroupements a dans certains cas été vérifiée par une analyse Adonis (Permutational Multivariate Analysis of Variance, 'Vegan', (Anderson 2001)), qui partitionne la variance entre différents classements.

La ressemblance entre répliqués d'un même échantillon d'eau a également été visualisée à l'aide d'une analyse factorielle (**CoA**), réalisée à l'aide du paquet R 'Ade4'. Cette analyse n'est toutefois pas une représentation des dissimilarités de Bray-Curtis, mais une projection des échantillons dans un plan depuis un espace dont chaque OTU constitue une dimension. Cette analyse permet de mettre en valeurs les plus grandes des différences entre échantillons.

Les dissimilarités de Bray-Curtis entre les échantillons des fractions de taille 0,2-5  $\mu\text{m}$  et 5-30  $\mu\text{m}$  de l'Etang des Vallées ont également été représentées par un **dendrogramme** montrant les **clusters** d'échantillons formés par average linkage à partir de la matrice de dissimilarités de Bray-Curtis ('Vegan').

D'autre part, nous avons dans certains cas voulu comparer la variabilité au sein d'un groupe d'échantillons avec celle au sein d'un autre groupe d'échantillons. Pour ce faire, les dissimilarités de Bray-Curtis entre toutes les paires d'échantillons de chacun des groupes sont représentées en **boîte à moustaches**, et la représentation des *notches* permet de déterminer si les médianes des deux distributions peuvent être considérées différentes. Cette représentation nous a par exemple permis de comparer la variabilité entre échantillons de sédiments avec la variabilité entre échantillons d'eau d'un même écosystème (chapitre 3).

Enfin, une dernière représentation a été utilisée afin de visualiser la saisonnalité de la dynamique des communautés. Pour ce faire, les dissimilarités de Bray-Curtis entre chaque librairie des échantillons de la première année d'échantillonnage et celle de l'échantillon prélevé 1 à 12 mois plus tard sont calculées. La moyenne, et l'erreur standard associée, des indices de dissimilarités entre échantillons séparés de 1, 2, 3,... et 12 mois sont calculées. Chaque moyenne est ainsi calculée sur 12 valeurs de dissimilarité entre échantillons séparés du même nombre de mois. Les dissimilarités moyennes entre échantillons séparés de 1 à 12 mois sont alors représentées dans un graphe.

## 2) Détermination du nombre d'OTU partagées par différents échantillons

Pour déterminer le nombre d'OTU partagées par différents groupes d'échantillons (e.g. issus du même écosystème, correspondant à la même fraction de taille de cellule), les nombres de séquences de toutes les librairies de chaque groupe ont été additionnés, et des diagrammes de Venn ont été construits à l'aide du paquet R 'gplots' (Bolker *et al.* 2012). Dans le chapitre 2, des *heatmaps* basées sur les données de présence/absence des OTU dans chaque écosystème (duplicats confondus) ont également été dessinées (paquet 'gplots') afin d'identifier les OTU partagées par un nombre donné d'écosystèmes.

## 3) Indices de richesse et de diversité

Ils ont été calculés à partir des comptages bruts du nombre de séquences par OTU et par échantillon. La richesse a été estimée par raréfaction. Cette technique permet de déterminer le nombre théorique d'OTU que l'on détecterait dans un sous-échantillon comprenant un nombre donné de séquences, ici le plus petit nombre de *reads* obtenu sur l'ensemble des librairies (Hurlbert 1971). Ainsi, l'estimation de la richesse d'un échantillon n'est pas fonction du nombre de séquences obtenues pour cet échantillon, et les richesses de différents échantillons peuvent être comparées. La diversité est estimée par l'indice de Simpson (Simpson 1949), qui correspond à la probabilité que deux séquences tirées au hasard appartiennent à deux OTU différentes. Il varie donc entre 0 et 1 et se calcule selon la formule suivante :  $S = 1 - \sum_i p_i^2$  avec  $p_i$  la fréquence de l'OTU<sub>i</sub> dans l'échantillon. Cet indice traduit à la fois la richesse en OTU et l'équiprobabilité des différentes OTU. L'équiprobabilité est également mesurée par l'indice d'équitabilité (Pielou 1966)  $e = - \sum_i p_i \cdot \ln(p_i) / \ln(S)$  où  $p_i$  est la fréquence de l'OTU<sub>i</sub> et S le nombre d'OTU observées dans l'échantillon considéré.

## 4) Détermination de l'abondance des OTU

Il est maintenant connu et reconnu depuis longtemps que les communautés microbiennes sont constituées d'un petit nombre d'OTU très abondantes qu'accompagne une multitude d'OTU beaucoup moins fréquentes (Pedrós-Alió 2006; Caron & Countway 2009b). Afin d'identifier les OTU dominantes dans les échantillons prélevés en avril 2012 (chapitre 2), nous avons calculé la proportion des *reads* de chaque OTU dans l'ensemble des *reads* des échantillons considérés. Par la suite, nous avons plutôt calculé pour chaque OTU une fréquence moyenne par échantillon. Cette seconde méthode a pour avantage de ne pas être



biaisée par les différences de nombres de séquences par librairie, plus élevées sur l'ensemble des échantillons du suivi temporel que sur ceux du mois d'avril 2012 uniquement.

### 5) Détermination de la fréquence des groupes fonctionnels

Afin d'avoir une première idée des rôles des petits eucaryotes dans le fonctionnement des écosystèmes d'eau douce peu profonds étudiés, nous avons grossièrement attribué une fonction potentielle aux organismes associés à chaque OTU, sur la base de leur affiliation taxonomique:

**Producteurs primaires (dont mixotrophes) :** Bacillariophyceae, Chlorophyta, Chrysophyceae-Synurophyceae, Cryptophyta, Dictyochophyceae, Dinophyta, Glaucophyta, Haptophyta, Rhodophyta, Streptophyta, Xanthophyceae.

**Parasites:** Apicomplexa, Rozellida-Cryptomycota, Ichthyosporea, Oomyceta, Perkinsea.

**Autres hétérotrophes putatifs :** Amoebozoa, Apusozoa, Bicosoecida, Centroheliozoa, Cercozoa, Choanoflagellida, Ciliata, Colpodellida, Discosea, Fungi, Katablepharida, Labyrinthulida, Malawimonadida, MAST, Metamonada, Metazoa, Telonema.

Aucune fonction n'a été attribuée aux OTU sans affiliation précise (i.e. autre alvéolé, autre amibe, autre opisthoconte, autre stramenopile ou affiliation inconnue).

Les fonctions étant attribuées sur la base d'affiliation à des groupes de rang taxonomique élevé, sans tenir compte de la diversité fonctionnelle au sein de ces groupes, elles sont peu précises. Les observations réalisées sur cette base donnent donc un premier aperçu du fonctionnement des assemblages de petits eucaryotes des mares et rus, mais resteront à vérifier par des études plus approfondies sur ce point.

### 6) Comparaison des caractéristiques physico-chimiques

Nous avons considéré un certain nombre de caractéristiques environnementales, susceptibles d'influencer la dynamique spatio-temporelle des OTU détectées: concentrations de certains nutriments, chlorophylle *a* et solutés totaux, température, pH, oxygénation de l'eau. Certaines concentrations en nutriments étaient inférieures à la concentration minimale détectable par les kits de dosage utilisés. Pour les analyses ultérieures, ces données manquantes ont été remplacées par la valeur minimale de détection du kit diminuée d'une valeur tirée dans une loi normale de moyenne zéro et de déviation standard  $10^{-6}$ , ceci afin de

ne pas considérer ces données comme des valeurs totalement manquantes, mais de prendre en compte le fait qu'elles étaient particulièrement faibles. En revanche, certaines valeurs étaient réellement manquantes (duplicats de dosages de chlorophylle *a* trop différents, sonde non fonctionnelle). Celles-ci ont été remplacées par la moyenne de toutes les autres valeurs du même paramètre dans les autres échantillons.

Afin de visualiser graphiquement les caractéristiques physico-chimiques associées aux différents échantillons, ainsi que la corrélation entre différents paramètres, des analyses en composante principale (ACP) ont été réalisées à l'aide du paquet R 'Ade4', sur des valeurs centrées et réduites puisque les différents paramètres étudiés correspondent à des grandeurs physiques différentes.

## 7) Identification de paramètres éventuellement impliqués dans la dynamique spatio-temporelle des communautés et des organismes qui les composent

Les corrélations éventuelles entre OTU et paramètres physico-chimiques ont été identifiées grâce à des analyses canoniques de correspondance (CCA, paquet R 'Ade4'). Dans le chapitre 2, toutes les OTU détectées dans les échantillons prélevés en avril 2012 et correspondant aux cellules recueillies sur les filtres de 0,2 µm de porosité sont prises en compte. Les fréquences des différentes OTU sont utilisées sans standardisation. Dans le chapitre 3, seules les OTU détectées dans au moins 25% des échantillons, et ayant une fréquence moyenne par échantillon supérieure ou égale à 1% sont prises en compte dans la CCA, et les fréquences des OTU sont soumises à une standardisation de Wisconsin. Pour les données de physico-chimie, les valeurs sous le seuil de détection des kits de dosages de nutriments ainsi que les données manquantes sont traitées comme pour l'ACP.

Nous avons voulu tester la corrélation éventuelle entre les dissimilarités de Bray-Curtis entre communautés et les distances géographiques entre écosystèmes ou les distances euclidiennes entre jeux de paramètres physico-chimiques. Les distances géographiques entre écosystèmes correspondent à des distances à vol d'oiseau estimées à partir des coordonnées GPS des sites ([http://biodiversityinformatics.amnh.org/open\\_source/gdmg/](http://biodiversityinformatics.amnh.org/open_source/gdmg/)), en utilisant le rayon moyen de la terre (Moritz 2000) comme sphéroïde. Au cours de l'étude des communautés prélevées en avril 2012 (chapitre 2), la corrélation entre la matrice de dissimilarité de Bray-Curtis et la matrice de distance géographique d'une part, et celle de distance entre les paramètres physico-chimiques d'autre part, a été testée grâce à un **test de Mantel** ('Vegan'). Dans l'analyse du suivi temporel des communautés (chapitre 3), la part de

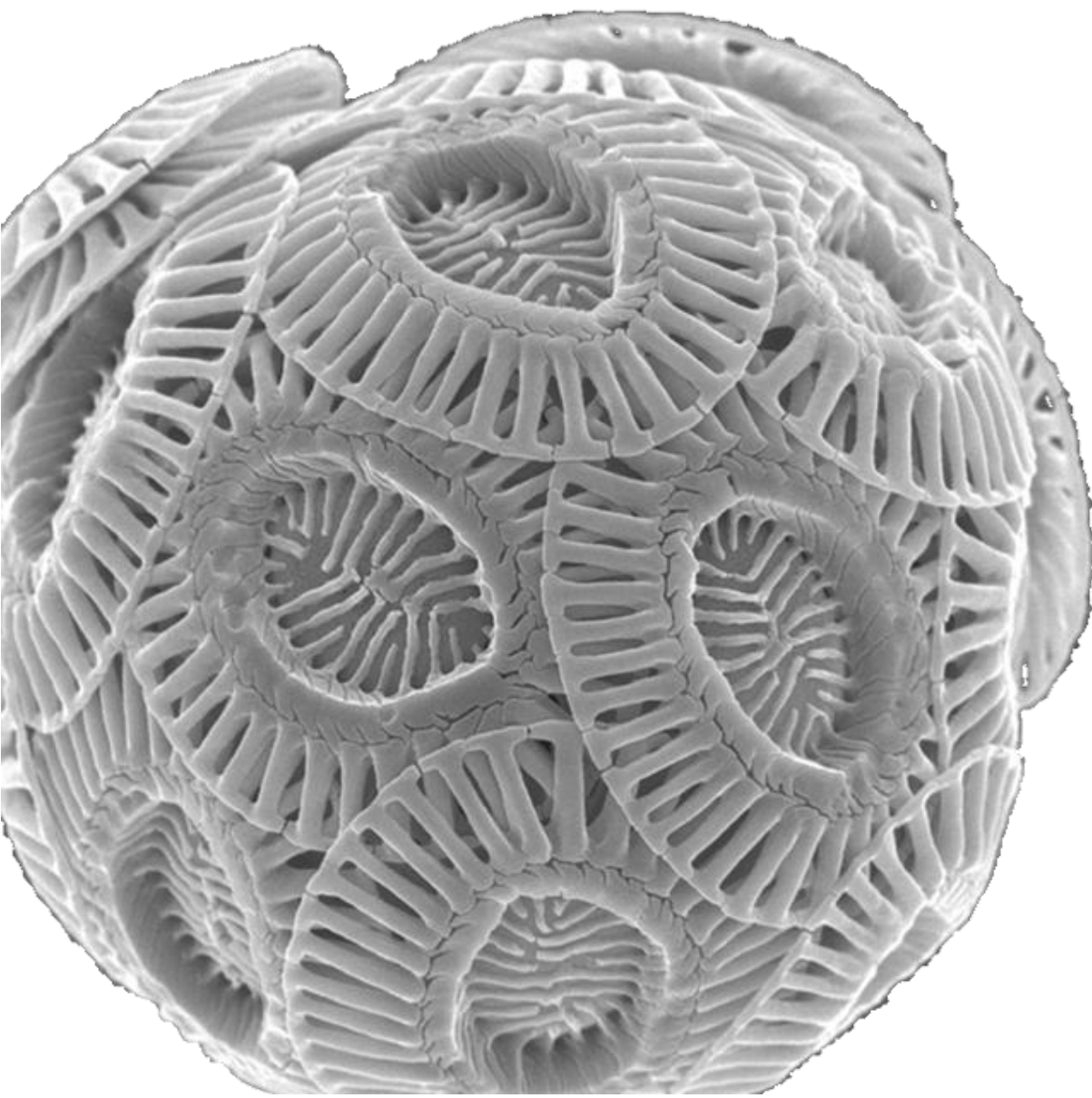
variance entre échantillons expliquée par les paramètres physico-chimiques est estimée par le rapport entre la somme des valeurs propres de la CCA et celle d'une analyse factorielle de correspondance (CoA) réalisée sur les mêmes proportions standardisées des mêmes OTU (Borcard *et al.* 1992).

## VIII. Reconstructions phylogénétiques

Outre les séquences issues de ce travail, les reconstructions phylogénétiques incluent des séquences représentatives de tous les taxa d'haptophytes ainsi que des séquences environnementales couvrant la quasi-totalité de l'ADNr 18S (chapitre 1), ou bien les premières séquences obtenues par BLAST de nos séquences et les séquences des organismes cultivés les plus proches (chapitre 2). Dans le cas de l'arbre des groupes de MAST, l'affiliation de nos séquences à un ribogroupe précis était particulièrement cruciale. En effet, la plupart des ribogroupes MAST sont actuellement composés uniquement de séquences marines (Massana *et al.* 2013). L'affiliation de nos séquences d'eau douce à certains de ces groupes exclusivement marins constituait donc une nouveauté (Cf. chapitre 2), et devait être d'autant plus fiable. Les reconstructions phylogénétiques des groupes de MAST incluaient donc également des représentants de tous les ribogroupes définis par Massana et collègues en 2013. Les séquences de références utilisées avec les *reads* issus du pyroséquençage ont été coupées au niveau de la séquence des amorces de PCR. Tous les alignements ont été réalisés avec le logiciel probcons (Do *et al.* 2005). Les positions informatives ont été sélectionnées dans les alignements à l'aide du logiciel Gblocks (Castresana 2000), en autorisant la conservation de petits blocs (séquences longues issues du clonage / séquençage Sanger, chapitre 1) ou en sélectionnant les options les plus permissives afin de conserver suffisamment de positions (séquences courtes issues du pyroséquençage, chapitre 2). Les arbres d'haptophytes (chapitre 1) ont été construits par maximum de vraisemblance à l'aide de TreeFinder (Jobb *et al.* 2004), en utilisant le modèle de substitution GTR, et une distribution du taux d'évolution des sites selon une loi gamma avec des sites invariants. Les arbres contenant les pyroséquences avaient pour objectif principal de déterminer de façon précise l'affiliation taxonomique des séquences, plutôt que d'identifier les relations phylogénétiques entre groupes. Ils ont donc été construits par une méthode plus rapide, ce qui était utile en raison du grand nombre de séquences impliquées. Ces reconstructions phylogénétiques ont été réalisées par approximation de maximum de vraisemblance, à l'aide du programme FastTree (Price *et al.* 2010). Tous les arbres ont été visualisés à l'aide du logiciel FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

# Chapitre 1 : Diversité des haptophytes en eau douce

---





## I. Introduction : résumé du contexte et objectifs

Les haptophytes constituent un groupe d'eucaryotes unicellulaires majeur dans les océans, tant quantitativement que fonctionnellement. Les nombreux travaux concernant ces organismes ont en effet démontré le rôle non négligeable qu'ils jouent dans les réseaux trophiques marins, via une forte production primaire (Jardillier *et al.* 2010), une bactériovorie importante (Unrein *et al.* 2014) et la production de toxines (Edvardsen & Imai 2006).

Cependant, très peu d'études ont cherché à décrire la diversité des haptophytes en eau douce. Pourtant, on sait qu'ils sont présents, parfois en forte abondance notamment lors de la formation d'efflorescences. Ils y jouent potentiellement un rôle écologique important ; plusieurs espèces toxiques sont capables d'entraîner la mort massive des poissons qu'elles côtoient. Enfin, un groupe jusqu'ici non observé en milieu marin et de position phylogénétique particulièrement intéressante a été détecté dans des mares ou lacs (Šlapeta *et al.* 2005; Shalchian-Tabrizi *et al.* 2011). Il nous est donc paru particulièrement intéressant d'explorer la diversité des haptophytes dans une large variété d'écosystèmes d'eau douce, depuis des petites mares peu profondes jusqu'à de grands lacs. Afin de comparer la diversité méconnue en eau douce avec celle beaucoup plus étudiée en milieu marin, mais également de valider la spécificité des nouvelles amorces que nous avons dessinées, deux colonnes d'eau marines (Atlantique sud et mer de Marmara) ont été intégrées à l'étude.

Les travaux réalisés en milieu marin ont révélé une forte sous-représentation des haptophytes dans les banques de séquences d'ADNr 18S réalisées à l'aide d'amorces de PCR générales des eucaryotes, par rapport à leur abondance estimée par d'autres méthodes (Liu *et al.* 2009). Nous avons donc voulu étudier dans ce travail la diversité des haptophytes à l'aide de deux méthodes moléculaires qui ont fait leur preuve pour d'autres taxa. L'amplification d'ADNr 18S à l'aide d'amorces *spécifiques* du groupe, avant clonage et séquençage Sanger des amplicons, a déjà fait ses preuves pour différents taxa tels que les diplonémides (Excavata, Lara *et al.*, 2009) ou les télonémides (Bråte *et al.* 2010a). C'est cette méthode que nous avons utilisée dans l'article "New haptophyte lineages and multiple independent colonisations of freshwater ecosystems" présenté dans ce chapitre. D'autre part, le pyroséquençage 454 permet d'obtenir un grand nombre de séquences, et ainsi de détecter des amplicons minoritaires. On peut donc potentiellement détecter des organismes réellement minoritaires dans l'échantillon, ainsi que les amplicons artefactuellement minoritaires après amplification par des amorces générales d'eucaryotes. Les résultats obtenus par cette seconde méthode sont exposés à la fin du présent chapitre.

## II.      **Manuscrit de l'article 1 : " New haptophyte lineages and multiple independent colonisations of freshwater ecosystems"**

*Environmental Microbiology Reports, 2013, Volume 5, Issue 2, Pages 322-332*

**New haptophyte lineages and multiple independent colonisations of freshwater ecosystems**

**Marianne Simon, Purificación López-García, David Moreira, and Ludwig Jardillier**

Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079 Université Paris-Sud, 91405 Orsay, France

Corresponding author: ludwig.jardillier@u-psud.fr

Running title: *New insights into haptophyte diversity*

**Keywords:** 18S rRNA, prymnesiophytes, pavlovophytes, specific primers, ecological transition



## A- Summary

The diversity and ecological relevance of small haptophytes in marine systems is increasingly recognized. Similar investigations in freshwater remain scarce, despite some recent studies showing the existence of divergent haptophyte lineages and indicating that these microalgae can occur at high abundance in lakes. We studied the diversity of haptophytes in a wide variety of marine, salty continental and, most particularly, freshwater environments by amplifying, cloning and sequencing 18S rRNA genes. For this purpose, we designed two sets of primers specific for the two recognized haptophyte classes, Prymnesiophyceae and Pavlovophyceae. We detected pavlovophyte sequences only in freshwater systems as well as several novel prymnesiophyte phylotypes both in freshwater and marine environment. In addition, we retrieved a cluster of sequences (HAP-3) from the Marmara Sea branching deeply in the haptophyte tree with no clear affiliation to either of the two recognized classes. Five of the freshwater prymnesiophyte phylotypes detected formed a divergent monophyletic group (EV) without close described representatives that branched within the Isochrysidales, a group of generally marine and most often calcifying coccolithophorids. The presence of several sequences of freshwater haptophytes scattered among marine taxa in phylogenetic trees, confirms the occurrence of several independent haptophyte transitions between marine and freshwater environments.

## B- Introduction

Haptophytes are unicellular aquatic, mostly marine, photosynthetic eukaryotes. They are broadly distributed and recognized as key players in biogeochemical cycles in marine ecosystems. Their wide distribution and relative high abundance in marine waters was initially shown by high performance liquid chromatography analyses of its characteristic pigment 19'-hexanoyloxyfucoxanthin (19-Hex) (Andersen *et al.* 1996). According to recent molecular investigations, haptophytes contributed from 30 to 50% of the photosynthetic standing stocks in the photic layer across the oceans in the year 2000 (Liu *et al.* 2009). Similarly, targeted metagenomic analyses suggest that haptophytes constitute on average 25% of global eukaryotic picophytoplankton carbon biomass (Cuvelier *et al.* 2010). Haptophytes may play a significant role in primary production even when they are not dominant; for instance, a new group of haptophytes was recently shown to contribute significantly to CO<sub>2</sub> fixation despite its low relative abundance (Jardillier *et al.* 2010). This might be due to their higher growth rates and their bigger size as compared to other more abundant planktonic phototrophs (Cuvelier *et al.* 2010) and, perhaps, also to the presence of efficient carbon concentrating mechanisms (Reinfelder 2011). Haptophytes may also play additional roles in the C-cycle through heterotrophic pathways, since some haptophytes have been shown to be mixotrophic (Legrand *et al.* 2001; Frias-Lopez *et al.* 2009).

In addition to their increasingly recognized ecological importance in oceans, haptophytes are at the heart of a phylogenetic debate, since their position in the eukaryotic tree remains unresolved. Haptophytes were initially affiliated to the chromalveolates, a eukaryotic super-group including also alveolates, stramenopiles (heterokonts) and cryptophytes, the ancestor of which was thought to have acquired a red algal plastid as secondary endosymbiont (for review, see (Keeling 2009)). However, whereas the monophyly of alveolates and stramenopiles is easily retrieved, recent molecular phylogenetic analyses suggest that cryptophytes and haptophytes do not form a monophyletic group with them. They rather seem to be sister to other clades, such as the kathablepharids, telonemids and centroheliid heliozoa, which do not have known photosynthetic members, forming another super-group recently named Hacrobia (Okamoto *et al.* 2009) or CCTH group (cryptophytes, centroheliid heliozoa, telonemids, haptophytes) (Burki *et al.* 2009), even if its monophyly remains discussed (Burki *et al.* 2012). At a finer phylogenetic scale, several questions also remain open. Haptophytes currently encompass two classes, the Pavlovophyceae and Prymnesiophyceae. The monophyly of those classes is supported by both morphology and molecular data, though the



phylogenetic relationships within each class are often reshuffled (see (Edwardsen & Medlin 2007; Edwardsen *et al.* 2011; Bendif *et al.* 2011) for recent propositions). Pavlovophytes have mainly been observed in littoral and brackish environments, though they have also been detected in freshwater systems by amplification of their 18S rRNA genes using pavlovophyte-specific primers (Shalchian-Tabrizi *et al.* 2011). They are currently represented by only four genera (namely *Diacronema*, *Exanthemachrysis*, *Pavlova* and *Rebecca*) distributed in four clades according to phylogenies based on 18S rRNA genes (Bendif *et al.* 2011). Compared to pavlovophytes, prymnesiophytes display a higher diversity and abundance and, consequently, have been studied more thoroughly. They comprise the orders Coccolithales, Isochrysidales, Phaeocystales, Prymnesiales, Syracosphaerales and Crepidolithales (de Vargas *et al.* 2007) or Zygodiscales (Jordan *et al.* 2004; Edwardsen & Medlin 2007) depending on the authors. The legitimacy of the orders Syracosphaerales and Crepidolithales or Zygodiscales is under discussion; they might be included within the Coccolithales (Edwardsen & Medlin 2007). Several haptophyte genera based on cell morphology later revealed to be polyphyletic based on molecular analyses (Edwardsen *et al.* 2011) and, conversely, several morphological species were found indistinguishable based on molecular markers (Bendif *et al.* 2011), which urges for a revision of haptophyte systematics based on reliable molecular phylogenies.

In spite of their ecological and phylogenetic importance, the diversity of haptophytes is not fully explored. Since the beginning of the century, the use of molecular methods based on the amplification, cloning and sequencing of 18S rRNA genes has uncovered a vast diversity of marine protist lineages affiliating to known taxa and a few undescribed divergent lineages scattered in the eukaryotic tree (Moon-van der Staay *et al.* 2001; López-García *et al.* 2001; Massana & Pedrós-Alió 2008). However, in those studies using general eukaryotic primers, the diversity and novelty of haptophytes was rather low. In the particular case of haptophytes, Moon-Van der Staay and colleagues (Moon-van der Staay *et al.* 2000) showed a great discrepancy between the proportion of haptophytes in marine picoplankton based on their characteristic 19-Hex pigment and their relative representation in 18S rDNA libraries. It is well known that the use of lineage-specific primers may enhance the recovery of diverse environmental sequences within a given taxon, as has been shown, for example, for cercozoans (Bass & Cavalier-Smith 2004) or diplomonads (Lara *et al.* 2009). Actually, the use of specific primers targeting the 28S rRNA gene allowed the discovery of an unsuspected diversity of non-calcifying prymnesiophytes in oceanic waters (Liu *et al.* 2009). The enrichment of haptophyte fractions by flow cytometry has been another alternative showing a hitherto unveiled diversity within the group. In this way, a novel clade of highly divergent

haptophytes was discovered in South Pacific waters (Shi *et al.* 2009). Subsequent metagenomic studies on haptophyte fractions sorted by flow cytometry confirmed a wide prymnesiophyte diversity in oceans (Cuvelier *et al.* 2010).

Compared to marine ecosystems, haptophytes in freshwater systems have been less studied and seem less diverse. Only a dozen freshwater haptophyte species have been described whereas over 400 species have been defined from marine environments (Preisig 2002). Most research efforts have concentrated on toxic species, since they trigger massive fish kills (Hansen *et al.* 1994; Edvardsen & Imai 2006). Described freshwater haptophytes belong to the orders Pavloales, Coccolithales and Prymnesiales (Preisig 2002), with *Chrysochromulina* and *Prymnesium* being the most common genera. Consistent with classical observations, molecular diversity studies of small eukaryotes in lakes (Richards *et al.* 2005; Lepère *et al.* 2008; Triadó-Margarit & Casamayor 2012) revealed neither a broad diversity nor a high proportion of haptophyte sequences; most of the sequences retrieved being close to *Chrysochromulina parva*. Nonetheless, under certain conditions, lake haptophytes can reach relatively high proportions in the euphotic zone. For instance, haptophytes accounted for 7.7 % of microbial eukaryotes in lake Aydat and up to 62.8% of small planktonic protists in Lake Bourget (Lepère *et al.* 2010). Most studies on freshwater haptophytes have concentrated in lakes. However, freshwater environments are varied and highly heterogeneous, and the eukaryotic diversity in these ecosystems is far from being described, especially in small freshwater bodies. In one of the first protist molecular surveys carried out in freshwater systems, Šlapeta *et al.* (2005) studied two ponds with different redox status and retrieved 18S rDNA phylotypes that formed a monophyletic group with pavlovophytes and prymnesiophytes but branched deeply in the haptophyte clade. The occurrence of this group (HAP-1) in freshwater systems has been confirmed by subsequent studies in Scandinavian lakes using 454 pyrosequencing of the V4 region of 18S rRNA genes (Shalchian-Tabrizi *et al.* 2011). This same study also suggested, based on the premise of a marine origin for haptophytes, that several transitions between marine and freshwater systems have occurred along the evolutionary history of the group. However, since the diversity and distribution of haptophytes in different ecosystem across a salinity gradient is not well known, this hypothesis awaits confirmation.

To contribute to a comprehensive account of haptophyte diversity, we designed new specific 18S rDNA primers covering the diversity of known prymnesiophytes, pavlovophytes and the rest of known haptophyte environmental lineages. We then carried out molecular surveys in a variety of ecosystems across a salinity gradient, including oceanic samples, a

variety of freshwater lakes and ponds, as well as brackish and hypersaline shallow lakes. Our study confirms a wider diversity of haptophytes in marine, as compared to freshwater, ecosystems, and reveals the occurrence of novel divergent lineages both in marine and freshwater systems. The ecological distribution of the new haptophyte lineages further supports the hypothesis of multiple transitions from marine to freshwater systems.

## C- Results and discussion

### *Prymnesiophyte and pavlovophyte diversity and distribution*

In order to study haptophyte diversity in a variety of ecosystems, we first designed specific 18S rDNA primers targeting separately Prymnesiophyceae and Pavlovophyceae plus other divergent lineages (see supplementary Materials and Methods). Specific haptophyte primers have been used in previous studies. However, they targeted either the 28S rRNA gene (Liu *et al.* 2009), for which a comprehensive species and environmental reference sampling is missing, or a particular haptophyte subset (pavlovophytes) (Shalchian-Tabrizi *et al.* 2011).

Location	GPS coordinates	Ecosystem type	Sample name	Size fraction
Charca Verde, campus University Paris-Sud	48°42'02"N 2°10'28"E	Freshwater pond	CV1 MVSF	0.22–5 mm n.d.
Chevreuse, PNR Haute Vallée de Chevreuse	48°42'18.6"N 2°02'23.5"E	Freshwater urban pond	CH1 CH2	0.22–5 mm > 5 mm
Etang de la Tour, PNR Haute Vallée de Chevreuse	48°39'39.2"N 1°52'48.3"E	Freshwater pond	To1 To2	5–30 mm 0.22–5 mm
Etang des Vallées, PNR Haute Vallée de Chevreuse	48°41'23.0"N 1°54'59.2"E	Freshwater shallow lake	EV2 EV3 EV6 EV7 EV8 EV9 EV10	0.22–5 mm 5–30 mm 0.22–5 mm 5–30 mm 0.22–5 mm 5–30 mm 0.22–5 mm
Mare Gabard, PNR Haute Vallée de Chevreuse	48°39'15.83"N 1°55'20.26"E	Freshwater pond, forest	MG1 MG2	0.22–5 mm 5–30 mm
Saint Robert, PNR Haute Vallée de Chevreuse	48°39'54.82"N 1°56'45.28"E	Freshwater village pond	SR1 SR2	0.22–5 mm 5–30 mm
Ru Sainte Anne, PNR Haute Vallée de Chevreuse	48°36'45.91"N 1°58'16.61"E	Freshwater brook	RSA1 RSA2	0.22–5 mm 5–30 mm
La Claye, PNR Haute Vallée de Chevreuse	48°36'31.72"N 1°56'17.33"E	Freshwater pond, forest	LC1 LC2	0.22–5 mm 5–30 mm
Etang du Perray, PNR Haute Vallée de Chevreuse	48°41'49"N 1°51'37"E	Freshwater pond	Pe1	5–30 mm
Etang de Pourras, PNR Haute Vallée de Chevreuse	48°42'52"N 1°50'39"E	Freshwater pond	Po1	5–30 mm
Etang de Cernay, PNR Haute Vallée de Chevreuse	48°40'50"N 1°57'55"E	Freshwater pond	Ce1	5–30 mm
Lac du Bourget, Savoie	45°44'N 05°51'E	Freshwater lake	BG1 BG6	0.22–60 mm 0.22–60 mm
Lac d'Annecy Northern basin, Savoie	45°54'N 06°07'E	Freshwater lake	AN1 AN2 AN6	0.22–60 mm 0.22–60 mm 0.22–60 mm
Etang d'en haut, Paimpont, Britain	48°00'30.66"N 2°13'36.06"W	Freshwater pond	Ht1	0.22–5 mm
Etang du Châtenay, Paimpont, Britain	48°00'14.40"N 2°13'48.36"W	Freshwater pond	Châ1 Châ2	0.22–5 mm 5–30 mm
South Atlantic	56°18'57"S 57°39'45"E	Marine	DH122 DH123 DH125 DH129	0.22–5 mm 0.22–5 mm 0.22–5 mm 0.22–5 mm
Marmara Sea, Central basin	40°50'18.48"N 28°01'24.24"E	Marine	Ma101 Ma125 Ma130 Ma135	0.22–5 mm 0.22–5 mm 0.22–5 mm 0.22–5 mm
Salada Chiprana, Spain	41°14'30"N 0°10'50"W	Hypersaline	SCH1 SCH2 SCH3	n.a. 0.22–5 mm 0.22–5 mm
Ornithological Park of Teich, France	44°38'27.72"N 1°01'14.04"W	Brackish	CPT2 CPT3 CPT4	n.d. n.d. n.d.

n.d., not done; n.a., not applicable.

**Table 1. Major characteristics of the samples analysed in this study. Additional details are provided in Table S1.**



Using our more inclusive haptophyte primer sets, we then amplified 18S rRNA genes from a variety of samples including freshwater, marine and saline systems (Table 1) (see Supplementary Information for methodological details).

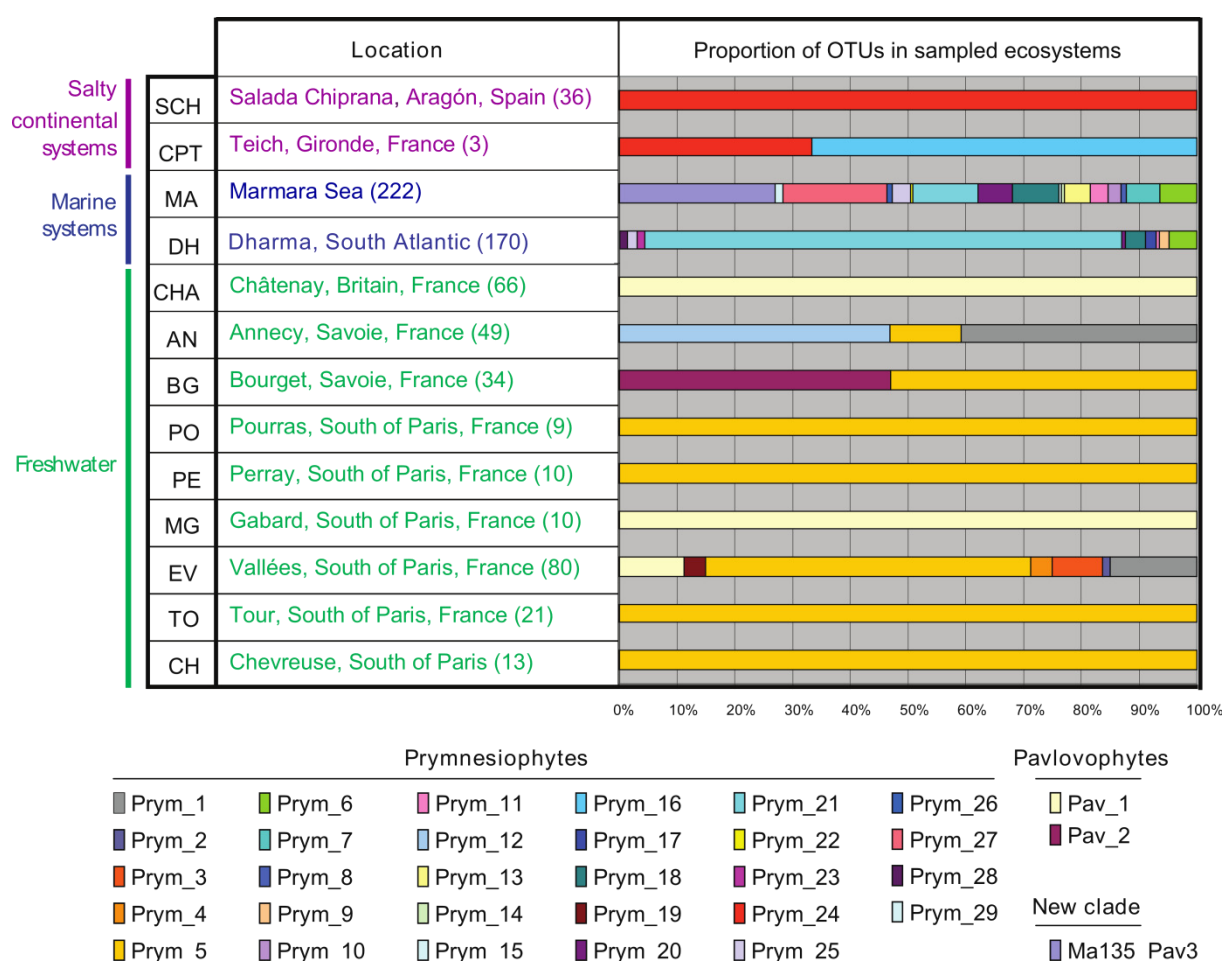
In all continental samples, haptophyte 18S rDNAs were amplified only after nested PCR reactions, suggesting that haptophytes were in low abundance in these ecosystems. Although we cannot totally discard a PCR bias, this is in agreement with the paucity of haptophyte sequences in most molecular surveys in lakes using general eukaryotic primers (Richards *et al.* 2005; Lepère *et al.* 2008; Triadó-Margarit & Casamayor 2012). This might suggest that several haptophyte lineages were rare at the sampling time, forming part of a 'seed bank' that may bloom or become more prevalent at different seasons. We identified a total of 32 haptophyte operational taxonomic units (OTUs, defined as groups of 18S rDNA sequences being more than 98% identical), most of them belonging to the prymnesiophytes, which were detected in the majority of the ecosystems sampled (Table S1). Most of the haptophyte diversity was observed in marine systems where 21 OTUs were retrieved out of the total of 32 detected, while only 9 and 2 OTUs were revealed in freshwater and salty (brackish or hypersaline) continental systems, respectively (Figs. 1, 2 and 3, Table S2). The detected phylotypes affiliated to the five prymnesiophyte orders generally accepted (de Vargas *et al.* 2007), namely Phaeocystales, Prymnesiales, Coccolithales, Isochrysidales and Syracosphaerales (Fig. 2), confirming the validity of the new prymnesiophyte primer set. In turn, pavlovophytes could only be detected in 4 out of 15 freshwater systems and were found neither in the continental saline waters nor in the marine environments sampled (although we did amplify a divergent set of sequences from marine samples using this set of primers; Table S1, see below). The two freshwater pavlovophyte OTUs retrieved in our study were affiliated to clade 4 (Bendif *et al.* 2011) (Fig. 3). Though the pavlovophyte diversity and distribution observed in our study were reduced, these results are consistent with current knowledge (Bendif *et al.* 2011).

### *New insights in the diversity of marine haptophytes*

More than half the 32 OTUs detected in the environments sampled were not closely related to described species (Table S2). Among them, five marine OTUs affiliated to previously described clades without any cultured representative species (Prym\_14, Prym\_20) or even to potential new clades (Prym\_13 and Prym\_29, Ma135\_Pav3) (Fig. 2 and 3, Table S2).



All sequences forming OTU Prym\_13 and Prym\_29 were detected in the Sea of Marmara. They grouped with the environmental sequence F01N5 that was retrieved at the epipelagic zone (0-200 m) in the Sargasso Sea (Not & Gausling 2007) and sequences SGZW1078 and FS04R14 from the euphotic zone (75 m depth) of Florida Straits and Sargasso Sea, respectively (Fig. 2 and Fig. S1) (Cuvelier *et al.* 2010). Interestingly, this group of sequences formed a new clade (that we call here clade B3) well nested within the Prymneshiales, along with clades B1 and B2 defined by Edvardsen *et al.* (2000, 2011). However, its position within the order is unstable; there is no significant statistical support for its sisterhood to either clade B1 or B2 (Fig. 2).



**Figure 1: Proportions of the OTUs detected in the different ecosystems investigated.** The number of sequenced haptophyte clones from each ecosystem is shown in brackets; the colour code utilized for the different OTUs is indicated at the bottom.

OTUs Prym\_14 and Prym\_20, affiliated to prymnesiophyte clade D and clade E respectively (Edvardsen *et al.* 2000, 2011), which are exclusively composed of marine



environmental sequences from the South East Pacific Ocean and the Equatorial Pacific Ocean (Edwardsen *et al.* 2000; Moon-van der Staay *et al.* 2000; Shi *et al.* 2009). Over the last decade, the phylogenetic position of those two clades D and E changed with the increasing number of sequences retrieved in marine environments (Edwardsen *et al.* 2000; Moon-van der Staay *et al.* 2000, 2001). Recently, Shi *et al.* (2009) proposed clade D to form an independent lineage at the base of all other prymnesiophyte orders within the haptophytes. In addition to its widespread distribution in marine environments, our study confirms that these sequences form an independent clade among prymnesiophytes, and should therefore represent a new order (Fig. 2). Shi *et al.* (2009) also proposed clade E to be a sister group of the coccolithophorid orders (Coccolithales, Isochrysidales, Syracosphaerales). Our phylogenetic analysis confirms the placement of the clade E within the clade formed by coccolithophorid orders, but without any clear sisterhood to any of the described orders (Fig. 2). The inclusion of new coccolithophorid sequences should help placing clade E within one of the described orders or support the erection of a novel order for organisms of this clade.

The use of pavlovophyte-specific primers revealed a divergent lineage, OTU Ma135\_Pav3, grouping 60 clone sequences (Table S2), only detected at surface (15 m depth) in the Sea of Marmara. The position of this divergent OTU is unstable. It branched either as a sister group of pavlovophytes, although with very low support (Fig. S2) or at the base of both pavlovophytes and prymnesiophytes (Fig. 3). The closest 18S rDNA sequence retrieved by a BLAST search against the Silva database (Pruesse *et al.* 2007) was retrieved from the South East Pacific Ocean (Shi *et al.* 2009) and shared only 93.9% identity with it (Table S2). Nonetheless, environmental sequences FS14K073 and EN360CTD001 (Cuvelier *et al.* 2010) and SHAX 513 (Orsi *et al.* 2012), which only partially overlapped our sequences, shared 97-98% identity with our sequence Ma135-Pav3-C1 and clustered together in phylogenetic trees (Fig. S2). This divergent OTU Ma135\_Pav3 did not branch with any of the other deeply divergent haptophyte lineages detected so far: the freshwater clade HAP-1 formed by the environmental sequences CV1-B1-97 and CV1-B2-32 retrieved from a suboxic pond (Šlapeta *et al.* 2005) and APB2H and AI9LL from Lake Finsevatn in Norway (Shalchian-Tabrizi *et al.* 2011) (Fig. 3 and Fig. S2) and the marine clade HAP-2 including sequences from the Biosope cruise T65.100 and T58.080 (Shi *et al.* 2009). Therefore, the newly detected clade, which we name here HAP-3, together with HAP-1 and HAP-2 might represent new class-level groups of haptophytes along with pavlovophytes and prymnesiophytes.

*Novel haptophytes in freshwater systems*

We detected a group of 5 OTUs unrelated to any described species or to any environmental sequence (Group EV, Fig. 2) in two freshwater lakes, the shallow lake Etang des Vallées and Lake Annecy (Table 1). Interestingly, this phylogenetic group affiliates to Isochrysidales, a coccolithophorid order that so far was thought to be composed exclusively of marine haptophytes. This order of haptophytes is composed of non-calcifying (*Isochrysis*) and calcifying (*Chrysotila*, *Emiliania*, *Gephyrocapsa*) genera. To our knowledge, *Hymenomonas roseola* (Coccolithales) is the only calcifying freshwater haptophyte that has been described (Manton *et al.* 1969). The lack of 18S rRNA gene sequences of *H. roseola* prevents us from establishing the phylogenetic relationship between this species and the sequences detected in our freshwater systems. Nonetheless, there is little chance that the EV group, affiliated to Isochrysidales, turns out to be closely related to *H. roseola*, which has been classified within the Coccolithales based on classical morphological description. Future morphological identification of members of the group EV should show whether the former are actually calcifying or not.

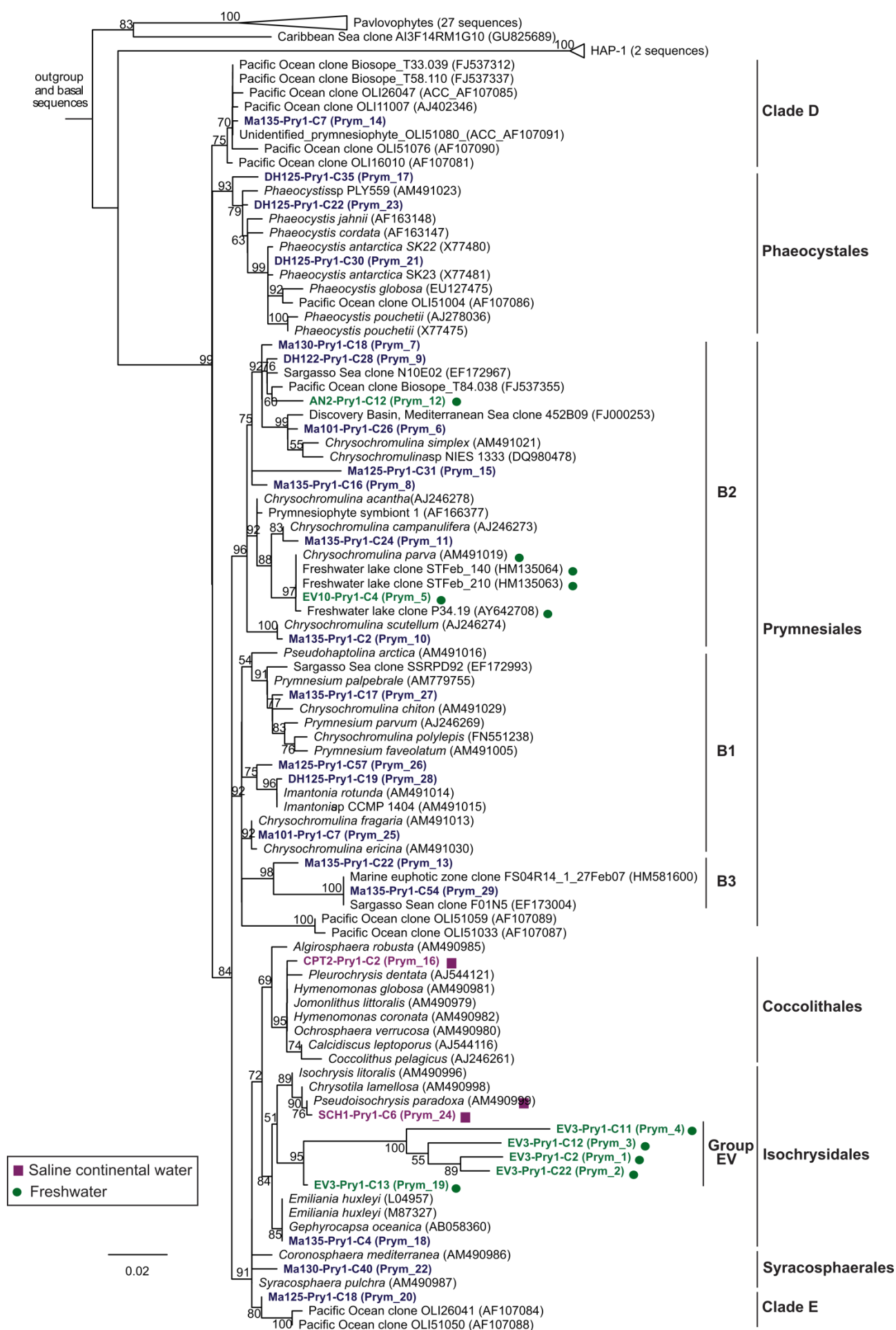
Even if our phylogenetic analysis strongly supports its monophyly, the Group EV is highly diverse. Indeed, sequences forming this group share 95% of identity only (average value calculated on 16 complete sequences). For comparison, 18S rRNA sequences of the polyphyletic genus *Chrysochromulina*, which is scattered in the order Prymnesiales (Fig. 2), share 96% identity (average value calculated on full-length 18S rDNA genes from 8 *Chrysochromulina* species by (Caron & Countway 2009a). Sequences belonging to group EV were found in both 0.22-5  $\mu\text{m}$  and 5-30  $\mu\text{m}$  fractions at the Etang des Vallées (Table 1), with OTUs Prym\_1 and Prym\_3 being shared by both the size fractions, although more diversity was retrieved in the biggest cell-size fractions. There are three possible explanations for the observation of members of this group in different fractions. One explanation would be imperfect size fractionation that could happen if, for instance, the cells are fragile and lyse during the filtration process, or can be deformed and passed through filters under the applied filtration pressure. Second, this could be explained by a diversity of cell sizes within members of Group EV due, for instance, to the presence of smaller gametes or different cell sizes in different life cycle phases (e.g. haploid-diploid transitions). Finally, a third possibility would be the non-obligate physical association of those organisms with bigger ones (e.g. symbiosis). The fact that OTUs Prym\_1, Prym\_2, Prym\_3 and Prym\_4 form a tight and diverse sub-group with much longer branches than most other prymnesiophyte sequences in our phylogenetic tree (Fig. 2) might suggest that they are mutualistic or parasitic haptophytes, since symbiotic

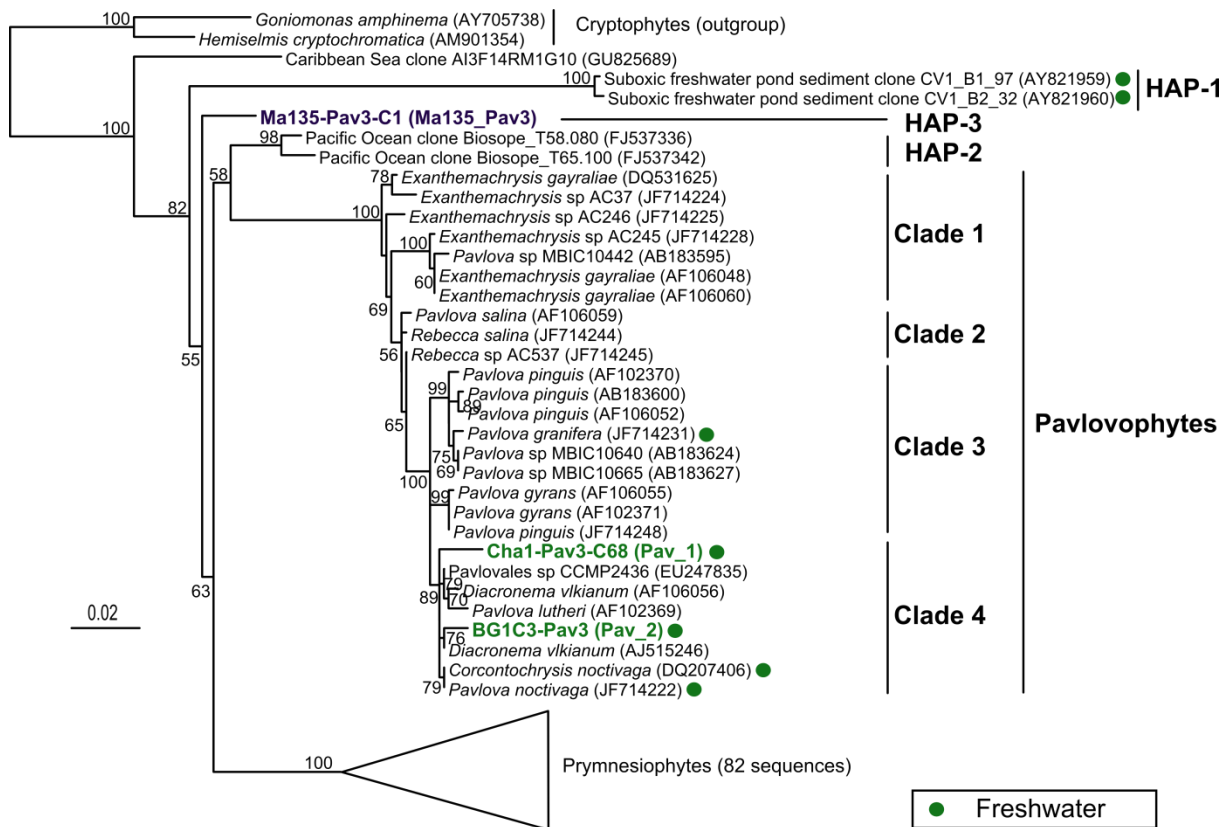
lineages tend to accelerate their evolutionary rate (Wernegreen 2002; Bromham 2009). Symbiotic prymnesiophytes have already been observed, for instance in association with the planktonic foraminifer *Globigerinella siphonifera* (Gast *et al.* 2000), which can either live with its host or be free-living.

This new Group EV has been detected in two geographically distant and ecologically different freshwater ecosystems: the shallow and oligo-mesotrophic shallow lake Etang des Vallées and the deep and oligotrophic Lake Annecy (Fig 1). Similarly, the divergent freshwater clade HAP-1 has also been detected in two different and geographically distant environments, first in the sediment of a suboxic pond in France (Šlapeta *et al.* 2005) then in the sediment of a high oligo- to mesotrophic alpine lake in Norway (Shalchian-Tabrizi *et al.* 2011). This suggests that, although not necessarily abundant, very divergent haptophyte lineages may still be found in the understudied freshwater systems. Investigating these lineages might help to reconstruct the evolutionary history of the group and to understand its ecology.

**Figure 2: (Next page) Maximum likelihood phylogenetic tree of 18S rDNA haptophyte sequences of marine, freshwater and salty continental habitats showing prymnesiophyte clades. A total of 758 non-ambiguously aligned positions were used to reconstruct the tree; gaps were excluded. Two cryptophyte sequences were used as outgroup. 18S rRNA gene sequences from this work are shown in bold. Pavlovophyte and HAP-1 branches are shown collapsed. Bootstrap values greater than 50% are shown at nodes. The scale bar represents the estimated number of substitutions per 100 positions per a unit length.**







**Figure 3: Maximum likelihood phylogenetic tree of 18S rDNA haptophyte sequences of marine, freshwater and salty continental habitats showing the diversity of pavlovophytes and basal haptophyte lineages.** A total of 851 non-ambiguously aligned positions were used to reconstruct the tree; gaps were excluded. 18S rRNA gene sequences from this work are shown in bold. The prymnesiophyte branch is shown collapsed. Bootstrap values greater than 50% are shown at nodes. The scale bar represents the estimated number of substitutions per 100 positions per unit branch length.

In addition to OTUs clustering within known haptophyte orders, the divergent lineage HAP-1 (Šlapeta *et al.* 2005; Shalchian-Tabrizi *et al.* 2011), together with the species *Pavlova granifera* (Green 1973) and six OTUs affiliated to the clade B1 of Prymnesiales (Finsevatt AI7UI, Finsevatt AKXPZ, Finsevatt AYOY0H) or to the pavlovophyte clade 4 (Finsevatt 8912, Svaersvann 14 and 16) (Shalchian-Tabrizi *et al.* 2011) have only been found in freshwater systems and might represent other examples of putative marine – freshwater transitions. Altogether, there might have been at least nine freshwater colonization events from marine waters. In addition, a few species, for which the 18S rRNA gene sequences are not yet available, have been visually observed in freshwater systems such as

*Chrysochromulina laurentiana*, *C. inornamenta*, *C. breviturrita* (Hansen *et al.* 1994; Nicholls 2003) and *Hymenomonas roseola* (Manton *et al.* 1969).

Continental salty ecosystems also harbour particular haptophytes as shown by OTUs Prym\_16 and Prym\_24 (Fig. 2) that were isolated in a brackish pond (France) and from Chiprana, belonging to the hypersaline lake complex in the central Ebro basin (Spain) (Jonkers *et al.* 2003). OTU Prym\_16 affiliated to *Jomonlithus littoralis* ALGO Je5, a coastal marine species, while OTU Prym\_24 was related to *Pseudoisochrysis paradoxa* CCAP949/1, isolated from the brackish York River Estuary in Virginia (USA) and *Chrysotila lamelosa*, a species often isolated from coastal marine regions (such as the strain ALGO HAP17) or brackish continental environments (such as CCAP 918/1 isolated in London, UK). Haptophytes recorded in salty continental systems thus appear to be phylogenetically close to species of either coastal habitats or continental brackish systems. This finding is in agreement with the importance of salinity as a barrier for marine – continental environments transitions, (Lozupone & Knight 2007; Logares *et al.* 2009), in spite of the fact that haptophytes seem to have crossed that barrier several times in the course of their evolution.

### Acknowledgements

We are most grateful to Stéphan Jacquet from the Institut National de la Recherche Agronomique, UMR42, CARRTEL, Thonon-les-Bains (France) and Anabel López-Archilla from the Universidad Autónoma de Madrid for generously providing samples of, respectively, lake Bourget (France) and Salada de Chiprana (Spain). We also thank P. Henry and N. Cagatay, chief scientists of the oceanographic cruise MARNAUT (2007), for providing P.L.G. the opportunity to collect samples in the Marmara Sea, as well as the Captain, the crew and the scientific team onboard of the R/V L'Atalante. We acknowledge the help of the Turkish Navy to protect our vessel in the zones of heavy ship traffic. We thank INSU for providing a CTD-rosette for water sampling and L. Fichen, who operated it. This work was financed by a CNRS-INSU program Ecosphère Continentale et Cotière (EC2CO). M.S. is recipient of a fellowship from the Ecole Normale Supérieure.



**New haptophyte lineages and multiple independent colonisations of freshwater ecosystems**

*Marianne Simon, Purificación López-García, David Moreira, and Ludwig Jardillier*

**D- Supplementary Information**

**Materials and methods**

*Study sites and sampling*

To look for potential new lineages of marine haptophytes, we used DNA of marine plankton from two very distant water columns in the South Atlantic and the Sea of Marmara collected, respectively, in December 1998 (DHARMA cruise) (López-García *et al.* 2001) and June 2007 (MARNAUT cruise (Lara *et al.* 2009) (Table 1 and Table S1). To explore a variety of ecosystems across a salinity gradient, sediment and water samples from a hypersaline lake (La Salada de Chiprana, in Spain, see (Jonkers *et al.* 2003) for complete description) were also studied. We also used DNA extracted from culture enrichments from a brackish pond in the ornithological park of Teich (France, close to the Arcachon laguna shores). In addition, fifteen geographically distant and close freshwater ponds, lakes and brooks were sampled in France from 2003 to 2011 at surface (Table 1 and Table S2).

Plankton cells of different size ranges (0.22-60  $\mu\text{m}$ , 5-30  $\mu\text{m}$  and 0.22-5  $\mu\text{m}$ ; see Table 1) were collected and concentrated either by filtering onto polycarbonate filters of different pore-sizes (Millipore), using Cell Traps (Mem-Teq) retaining cells ranging from 0.22 to 60  $\mu\text{m}$  in diameter thanks to a prefiltration through a 60  $\mu\text{m}$  pore-size filter, or by centrifugation after culture enrichment with K-medium (Keller *et al.* 1987) and Volvic mineral water. DNA was then extracted from the concentrated cells in a way depending on the collecting method used. Total DNA was extracted from filters stored at  $-20^{\circ}\text{C}$  using the PowerSoil DNA extraction kit (MoBio) according to the manufacturer's instructions. DNA was eluted in 50-80  $\mu\text{L}$  of Tris-HCl (10 mM pH 8, stock). Total DNA extraction from samples concentrated in CellTraps, flash frozen and stored at  $-80^{\circ}\text{C}$ , was performed according to a protocol adapted from Jardillier *et al.* (Jardillier *et al.* 2010). Briefly, 1  $\mu\text{L}$  SDS 10% and 1  $\mu\text{L}$  Proteinase K were added to 100  $\mu\text{L}$  culture. After 45 min incubation at  $50^{\circ}\text{C}$ , cells were incubated 10 min at room temperature with 4  $\mu\text{L}$  GenElute and 16  $\mu\text{L}$  Sodium Acetate. DNA was then precipitated and washed in ethanol before being resuspended in 10  $\mu\text{L}$  10 mM Tris. Finally,

when culture enrichments were performed, PCR were performed directly on culture pellets washed and resuspended on 10 mM Tris, or after DNA extraction as for DNA extraction from CellTraps.

#### *18S rRNA gene amplification, cloning and sequencing*

18S rRNA gene sequences were amplified using two sets of primers designed to target specifically Prymnesiophyceae (Pry421F: 5'-AGCAGGCGCGTAAATTGCCCCG-3' + Pry1572R: 5'-TCAACGYRCGCTGATGACA-3') and Pavlovophyceae (Hap220F: 5'-ACCGGTCTCCGGTTGCGTGC-3' + Pav1702R: 5'-TAGATGATAAGGTTTGGGTG-3') respectively. PCR amplifications were performed in a total reaction volume of 25 µl using 1 µL of DMSO (Sigma-Aldrich), 1.5 mM MgCl<sub>2</sub>, 0.2 mM dNTP, 0.2 µM of primers, 1 × PCR buffer and 0.5 U HotStart Taq polymerase (Taq Platinum, Invitrogen). TouchDown amplification conditions consisted of 10 cycles of 94°C for 15 s, 60 to 51°C for 30 s, 72°C for 2 min, followed by 25 cycles of 94°C for 15 s, 55°C for 30 to 54 s and 72°C for 2 min and a final extension step of 72°C for 7 min. 18S rRNA gene sequences from continental samples were amplified using a nested PCR, first using general Eukaryote primers 82F and 1498R as described in López-García *et al.* (López-García *et al.* 2003) and then using either prymnesiophytes or pavlovophytes specific primers as described above. PCR gene amplicons were cloned using the TOPO TA cloning kit (Invitrogen) according to the supplier's instructions. 20-40 clones per sample (Fig. 1 and Table S2) were chosen randomly and the 18S rDNAs contained in those clones were sequenced using the forward primer previously used for the DNA amplification (Beckman Coulter Genomics, Takeley, IK).

#### *Sequence and molecular phylogenetic analyses*

Sequences were manually screened for the presence of chimeric artefacts by independent BLAST searches of the 5' and the 3' halves as well as by using the software package KeyDNATool (<http://KeyDNATools.com>). The phylogenetic positions of the sequences were initially assessed by blasting them against the Silva SSU104-parc1000 database (Pruesse *et al.* 2007). Sequences affiliated to haptophytes were aligned with the ProbCons software (Do *et al.* 2005). Distance matrices calculated using ClustalX (Larkin *et al.* 2007) were used to construct Operational Taxonomic Units (OTUs) using a 98% sequence similarity cut-off with Mothur software (Schloss *et al.* 2009). Representative sequences of all OTUs were then sequenced using the reverse amplification primer in order to obtain complete clone sequences of about 1400 bp for pavlovophytes and 1070 bp for prymnesiophytes. For the final

phylogenetic trees, we incorporated representative sequences of all cultured haptophyte taxa, plus a wide representation of environmental sequences that were nearly full-length. We also reconstructed trees containing short environmental haptophyte sequences (<500 bp) with pivotal positions allowing for an important proportion of gaps (see Figs. S1 and S2 below). However, due to the limited number of non-ambiguously aligned positions available for phylogenetic analysis, short haptophyte sequences were not included in final phylogenetic trees where the position of divergent lineages was to be ascertained (Figs. 2 and 3). The number of positions selected from the multiple alignments was selected using Gblocks (Castresana, 2000) or BMGE (Criscuolo & Gribaldo 2010). Phylogenetic trees were reconstructed by maximum likelihood (ML) using Treefinder (Jobb et al., 2004) applying a GTR model of sequence evolution considering 4 rate categories and taking among-site rate variation into account by using a four-category discrete approximation of a  $\Gamma$  distribution. ML bootstrap proportions were inferred using 1,000 replicates. Phylogenetic trees were visualized using the program FIGTREE (<http://tree.bio.ed.ac.uk/software/figtree/>). Complete sequences affiliated to haptophytes have been deposited in GenBank under accession numbers JX680338 - JX680446.

*Table S1. Major characteristics of the samples analysed in this study. The positive or negative amplification of 18S rRNA genes with prymnesiophyte (Prym.) or pavlovophyte (Pav.)-specific primers is indicated with '+' or '-' signs. n.a., not applicable; n.d., not done.*

Location	GPS coordinates	Ecosystem type	Sample name	Sampling date	Depth	Size fraction	Volume filtered	Prym.	Pav.
Charca Verde, campus University Paris-Sud	48°42'02"N 2°10'28"E	Freshwater pond	CV1	November 2003	Surface	0.22 - 5 µm	150 ml	-	-
			MVSF	February 2005	Sediment surface	n.d.	n.a.	-	-
Chevreuse, PNR Haute Vallée de Chevreuse	48°42'18.6"N 2°02'23.5"E	Freshwater urban pond	CH1	December 2007	Surface	0.22 - 5 µm	500 ml	+	-
			CH2	December 2007	Surface	>5 µm	500 ml	+	-
Etang de la Tour, PNR Haute Vallée de Chevreuse	48°39'39.2"N 1°52'48.3"E	Freshwater pond	To1	June 2010	Surface	5 – 30 µm	500 ml	+	-
			To2	June 2010	Surface	0.22 - 5 µm	200 ml	+	-
Etang des Vallées, PNR Haute Vallée de Chevreuse	48°41'23.0"N 1°54'59.2"E	Freshwater shallow lake	EV2	February 2010	Surface	0.22 - 5 µm	300 ml	+	-
			EV3	February 2010	Surface	5 – 30 µm	300 ml	+	-
			EV6	June 2010	Surface	0.22 - 5 µm	100 ml	-	-
			EV7	June 2010	Surface	5 – 30 µm	500 ml	-	-
			EV8	September 2010	Surface	0.22 - 5 µm	140 ml	+	-
			EV9	September 2010	Surface	5 – 30 µm	200 ml	+	-
			EV10	March 2011	Surface	0.22 - 5 µm	100 ml	+	+
Mare Gabard, PNR Haute Vallée de Chevreuse	48°39'15.83"N 1°55'20.26"E	Freshwater pond, forest	MG1	March 2011	Surface	0.22 - 5 µm	100 ml	-	+
			MG2	March 2011	Surface	5 – 30 µm	200 ml	-	-
Saint Robert, PNR Haute Vallée de Chevreuse	48°39'54.82"N 1°56'45.28"E	Freshwater village pond	SR1	March 2011	Surface	0.22 - 5 µm	100 ml	-	-
			SR2	March 2011	Surface	5 – 30 µm	200 ml	-	-
Ru Sainte Anne, PNR Haute Vallée de Chevreuse	48°36'45.91"N 1°58'16.61"E	Freshwater brook	RSA1	March 2011	Surface	0.22 - 5 µm	100 ml	-	-
			RSA2	March 2011	Surface	5 – 30 µm	500 mL	-	-
La Claye, PNR Haute Vallée de Chevreuse	48°36'31.72"N 1°56'17.33"E	Freshwater pond, forest	LC1	March 2011	Surface	0.22 - 5 µm	100 ml	-	-
			LC2	March 2011	Surface	5 – 30 µm	500 ml	-	-
Etang du Perray, PNR Haute Vallée de Chevreuse	48°41'49"N 1°51'37"E	Freshwater pond	Pe1	June 2010	Surface	5 – 30 µm	200 ml	+	-
Etang de Pourras, PNR Haute Vallée de Chevreuse	48°42'52"N 1°50'39"E	Freshwater pond	Po1	June 2010	Surface	5 – 30 µm	200 ml	+	-
Etang de Cernay, PNR Haute Vallée de Chevreuse	48°40'50"N 1°57'55"E	Freshwater pond	Ce1	June 2010	Surface	5 – 30 µm	200 ml	-	-
Lac du Bourget, Savoie	45°44'N 05°51'E	Freshwater lake	BG1	May 2009	Surface	0.22 - 60 µm	10 l	n.d.	+
			BG6	April 2009	15 m	0.22 - 60 µm	10 l	+	-
Lac d'Annecy Northern basin, Savoie	45°54'N 06°07'E	Freshwater lake	AN1	May 2009	Surface	0.22 - 60 µm	10 l	-	-
			AN2	May 2009	15 m	0.22 - 60 µm	10 l	+	n.d.
			AN6	April 2009	15 m	0.22 - 60 µm	10 l	+	-
Etang d'en haut, Paimpont, Britain	48°00'30.66"N 2°13'36.06"W	Freshwater pond	Ht1	April 2010	Surface	0.22 - 5 µm	500 ml	-	-
Etang du Châtenay, Paimpont, Britain	48°00'14.40"N 2°13'48.36"W	Freshwater pond	Châ1	April 2010	Surface	0.22 - 5 µm	500 ml	-	+
			Châ2	April 2010	Surface	5 - 30 µm	200 ml	-	-
South Atlantic	56°18'57"S 57°39'45"E	Marine	DH122	December 1998	10 m	0.22 - 5 µm	23 l	+	-
			DH123	December 1998	25 m	0.22 - 5 µm	22 l	+	-
			DH125	December 1998	100 m	0.22 - 5 µm	20.5 l	+	-
			DH129	December 1998	2000 m	0.22 - 5 µm	32.5 l	+	-
Marmara Sea, Central basin	40°50'18.48"N 28°01'24.24"E	Marine	Ma101	June 2007	997 m	0.22 - 5 µm	10-15 l	+	-
			Ma125	June 2007	100 m	0.22 - 5 µm	5 l	+	-
			Ma130	June 2007	25 m	0.22 - 5 µm	4 l	+	-
			Ma135	June 2007	15 m	0.22 - 5 µm	2.5 l	+	+
Salada Chiprana, Spain	41°14' 30" N 0°10' 50" W	Hypersaline	SCH1	February 2010	Sediment surface	n.a.	n.a.	+	-
			SCH2	February 2010	0.5 m	0.22 - 5 µm	100 ml	+	-
			SCH3	February 2010	3.5 m	0.22 - 5 µm	25 ml	+	-
Ornithological Park of Teich, France	44°38'27.72"N 1°01'14.04"W	Brackish	CPT2	February 2010	Culture	n.d.	n.a.	+	-
			CPT3	February 2010	Culture	n.d.	n.a.	+	-
			CPT4	February 2011	Culture	n.d.	n.a.	+	-

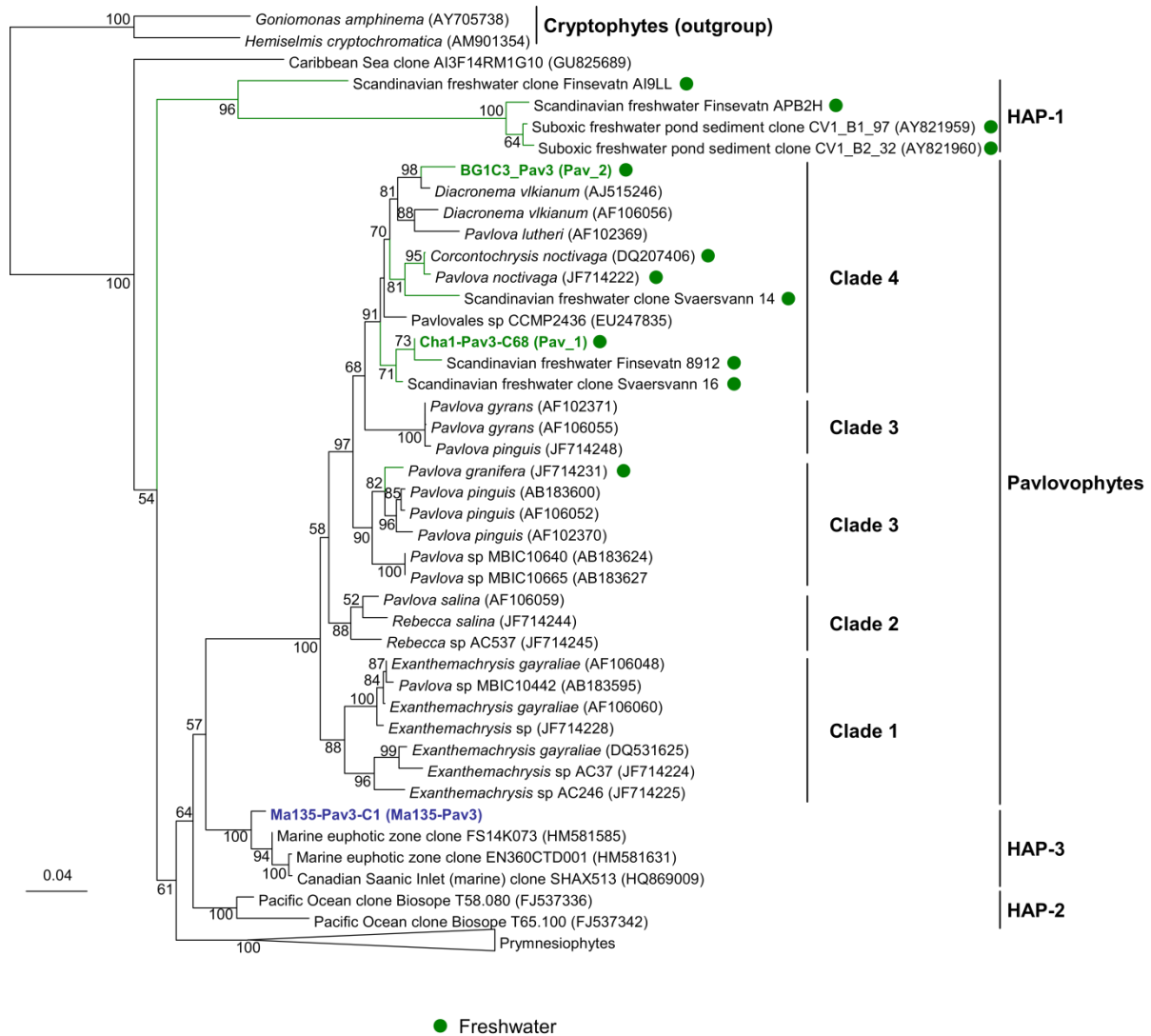
		Freshwater										Sea	Salt lakes		First hit (% identity)	
		Chèvreuse	Tour	Vallées	Gabard	Perray	Pourras	Bouget	Anney	Châtenay	Dharma	Marmara	Teich	Salada		Chiprana
OTU	Representative Sequence															
Ma135-Pav3	Ma135-Pav3-C1											60				Pacific Ocean clone Biosope T58.080 FJ537336 (93.9%)
Pav_1	Cha1-Pav3-C68			9	10					66						Corcontochrysis noctivaga DQ207406 (97.0%)
Pav_2	BG1C3-Pav3						16									Diacronema vikianum AJ515246 (98.4%)
Pym_1	EV3-Pry1-C2			12				20								Emiliana huxleyi M87327 (92.4%)
Pym_2	EV3-Pry1-C22			1												Emiliana huxleyi M87327 (97.2%)
Pym_3	EV3-Pry1-C12			7												Emiliana huxleyi M87327 (95.2%)
Pym_4	EV3-Pry1-C11			3												Emiliana huxleyi M87327 (92.6%)
Pym_5	EV10-Pry1-C4	13	21	45		10	9	18	6							Chrysochromulina parva AM491019 (99.5%)
Pym_6	Ma101-Pry1-C26										8	14				Mediterranean sea clone 452B09 FJ000253 (98.9%)
Pym_7	Ma130-Pry1-C18											13				Sargasso Sea clone N10E02 EF172967 (98.7%)
Pym_8	Ma135-Pry1-C16											2				Pymnesiophyte symbiont 1 AF166377 (98.5%)
Pym_9	DH122-Pry1-C28										3					Sargasso Sea clone N10E02 EF172967 (98.7%)
Pym_10	Ma135-Pry1-C2											5				Chrysochromulina scutellum AJ246274 (99.9%)
Pym_11	Ma135-Pry1-C24										1	7				Chrysochromulina campanulifera AJ246273 (99.2%)
Pym_12	AN2-Pry1-C12							23								Sargasso Sea clone N10E02 EF172967 (97.9%)
Pym_13	Ma135-Pry1-C22											10				Pymnesiophyte symbiont 1 AF166377 (96.7%)
Pym_14	Ma135-Pry1-C7											1				Pacific Ocean clone Biosope T58.110 FJ537337 (99.9%)
Pym_15	Ma125-Pry1-C31											1				Chrysochromulina scutellum AJ246274 (97.2%)
Pym_16	CPT2-Pry1-C2												2			Jomonolithus littoralis AM490979 (99.8%)
Pym_17	DH125-Pry1-C35										3					Phaeocystis antarctica SK22 X77480 (97.9%)
Pym_18	Ma135-Pry1-C4										6	18				Emiliana huxleyi M87327 (100%)
Pym_19	EV3-Pry1-C13			3												Emiliana huxleyi M87327 (98.4%)
Pym_20	Ma125-Pry1-C18										1	13				Syracosphaera pulchra AM490987 (99.0%)
Pym_21	DH125-Pry1-C30										145	25				Phaeocystis antarctica SK23 X77481 (99.9%)
Pym_22	Ma130-Pry1-C40											1				Syracosphaera pulchra AM490987 (98.7%)
Pym_23	DH125-Pry1-C22										1					Phaeocystis sp PLY559 AM491023 (97.5%)
Pym_24	SCH-Pry1-C6											1	1	36		Pseudoisochrysis paradoxa AM490999 (99.6%)
Pym_25	Ma101-Pry1-C7										1	7				Chrysochromulina fragaria AM491013 (99.9%)
Pym_26	Ma125-Pry1-C57											2				Chrysochromulina ericina AM491030 (97.3%)
Pym_27	Ma135-Pry1-C17											40				Sargasso Sea clone SSRPD92 EF172993 (99.1%)
Pym_28	DH125-Pry1-C19										1					Imantonia rotunda AM491014 (99.9%)
Pym_29	Ma135-Pry1-C54											3				Sargasso Sea clone F01N5 EF173004 (99.8%)

**Table S2. OTUs identified in this work.** The name of representative sequences for each OTU, their first BLAST hit in the Silva database SSU104, their percentage of similarity as well as the total number of sequences retrieved in each system are given.



**Figure. S1.** Maximum likelihood phylogenetic tree of 18S rDNA haptophyte sequences showing all prymnesiophyte OTUs retrieved in this study, their first hit by blast against the SILVA database SSU104 and sequences representing known orders and environmental lineages, including partially overlapping sequences to our sequences, especially including sequences retrieved from freshwaters. Two cryptophyte sequences were used as outgroup. The alignment contained 1682 selected position. Positions on a 2,147bp alignment having less than 50% gaps were retained to reconstruct the tree using BMGE. 18S rRNA gene sequences from this work are shown in bold; colour codes correspond to marine (blue), freshwater (green) or salty continental (purple) systems. Full circles indicate sequences of freshwater and salty continental habitats, other sequences are from marine ecosystems. Bootstrap values greater than 50% are shown at nodes (1000 replicates). The scale bar represents the number of substitutions per 100 positions per a unit branch length.





**Figure S2.** Maximum likelihood phylogenetic tree of 18S rDNA haptophyte sequences of marine, freshwater and salty continental habitats. This tree was built using the same sequences as in Fig. 3. plus shorter and/or partially overlapping environmental sequences related to our sequences. The 82 prymnesiophyte sequences used to construct the tree are shown collapsed. The alignment contained 1,678 selected positions (positions with less than 50% gaps were selected using BMGE on a 2,131 bp alignment). 18S rRNA gene sequences from this work are shown in bold. Full circles indicate freshwater sequences, other sequences are from marine ecosystems. Bootstrap values greater than 50% are shown at nodes (1000 replicates). The scale bar represents the number of substitutions per 100 positions per unit branch length.

### III. Epilogue : Exploration de la diversité des haptophytes dans quatre mares et un ru par pyroséquençage 454

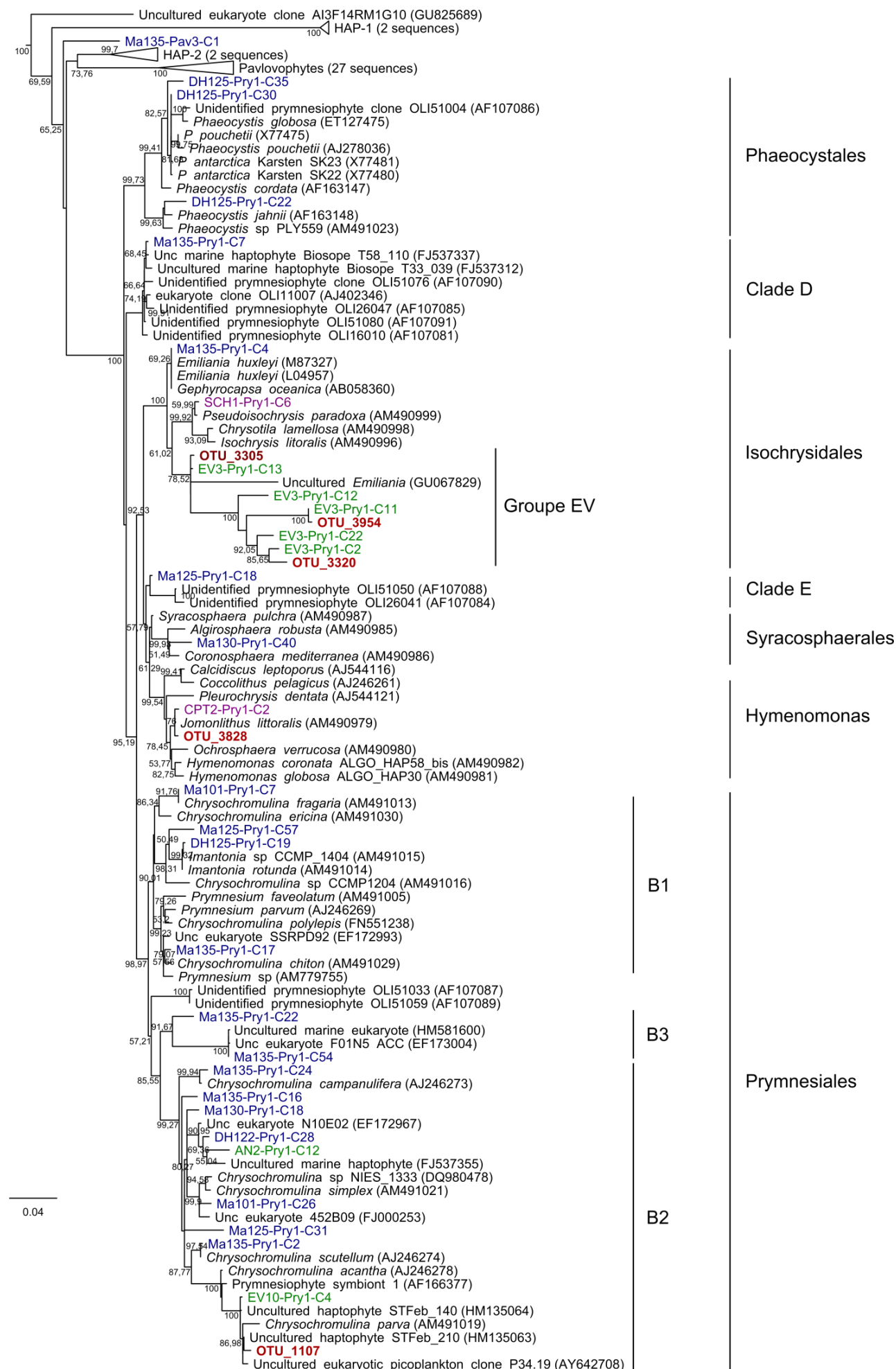
Comme ce sera expliqué dans les chapitres suivants, nous avons étudié la diversité générale des eucaryotes unicellulaires dans quatre mares et un ruisseau du Parc Naturel de la Haute Vallée de Chevreuse (PNR-HVC), au cours d'un suivi mensuel sur 2 ans (avril 2011 à avril 2013), par amplification à l'aide d'amorces générales d'eucaryotes puis pyroséquençage de fragments d'ADNr 18S (Cf. Matériel et Méthodes général).

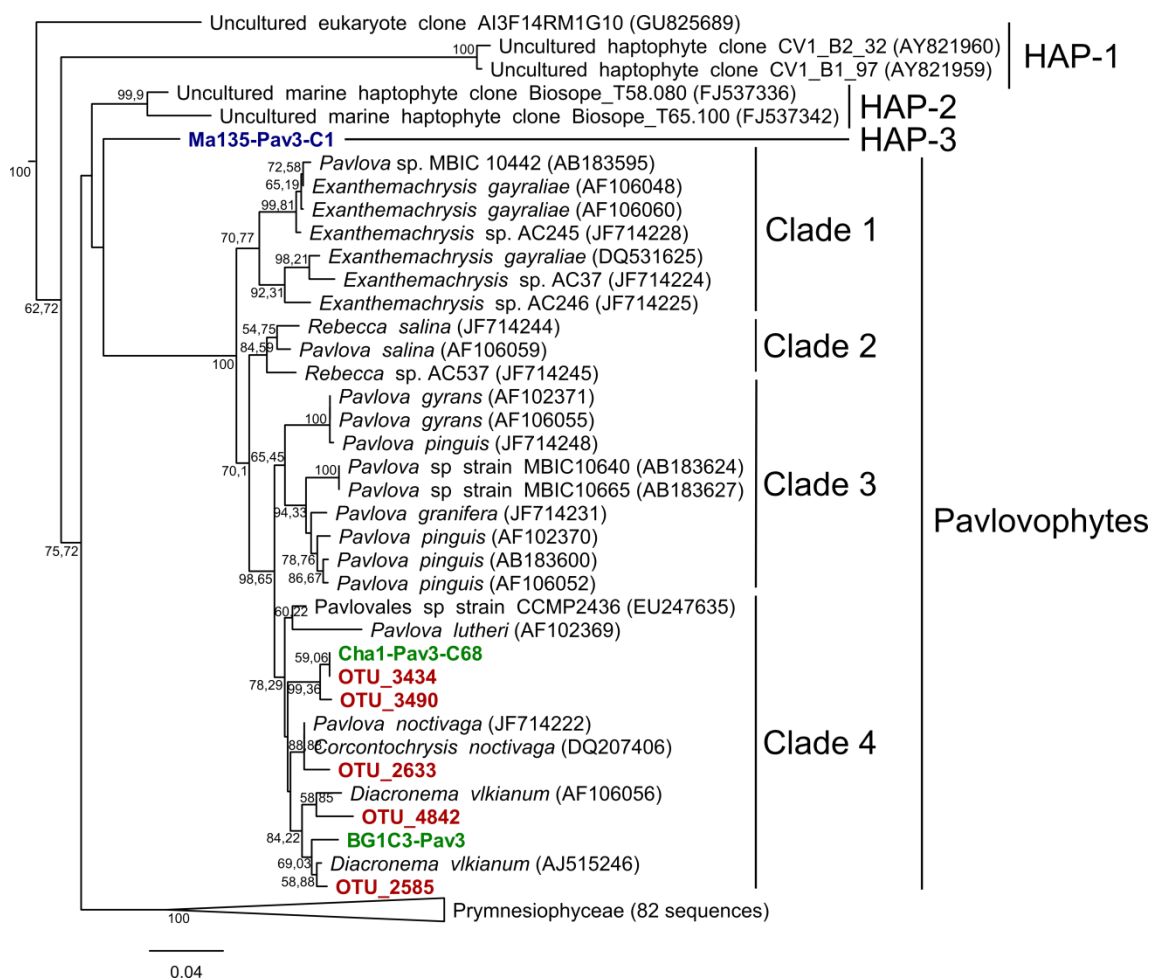
Seulement 10 OTU parmi les 4 455 détectées, tous échantillons confondus, ont été affiliées aux haptophytes d'après la similarité de leurs séquences représentatives avec des séquences d'haptophytes (BLAST). Par clonage et séquençage Sanger d'amplicons obtenus à l'aide d'amorce spécifiques des haptophytes, nous avons détecté 9 OTU différentes en eau douce (Table S2 du matériel supplémentaire de l'article). Ainsi, malgré la grande profondeur de séquençage permise par le pyroséquençage et le nombre relativement élevé d'échantillons d'eau douce analysés, nous n'avons pas détecté une plus grande richesse d'OTU haptophytes.

D'autre part, nous avons réalisé des reconstructions phylogénétiques reprenant les séquences utilisées dans les figures 2 et 3 de l'article et incluant les 5 séquences représentatives des OTU affiliées aux prymnésiphytes (Figure 4) et les 5 séquences des OTU affiliées aux pavlovophytes (Figure 5). Les mêmes procédés ont été utilisés, depuis l'alignement jusqu'à l'arbre final, la seule exception étant le choix des paramètres les plus permissifs de Gblocks (autorisation des gaps notamment) pour la sélection des positions informatives, afin que les séquences courtes obtenues par 454 ne réduisent pas drastiquement la longueur de l'alignement.

*Figure 4: Arbre phylogénétique en maximum de vraisemblance des séquences d'haptophytes issues des clonages / séquençage Sanger (en vert et bleu) et pyroséquençage 454 (en rouge), montrant la diversité des prymnésiphytes. 1 610 positions ont été utilisées pour reconstruire l'arbre. 2 séquences de cryptophytes servent de groupe externe (non montrées).*

# Chapitre 1



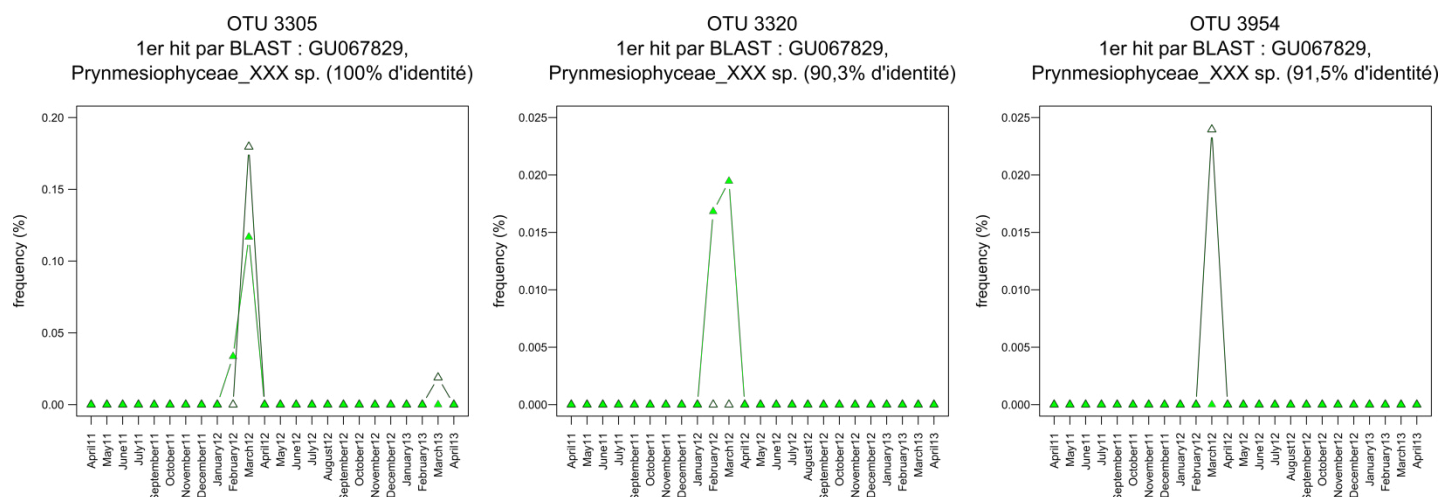


**Figure 5:** Arbre phylogénétique en maximum de vraisemblance des séquences d'haptophytes issues des clonages / séquençage Sanger (en vert et bleu) et pyroséquençage 454 (en rouge), montrant la diversité des pavlovophytes et des lignées d'haptophytes basaux. 1 596 positions ont été utilisées pour reconstruire l'arbre. 2 séquences de cryptophytes servent de groupe externe (non montrées).

Outre leur faible nombre, les OTU d'haptophytes détectées par pyroséquençage sont toutes proches d'OTU déjà détectées en eau douce ou dans un écosystème continental salé à l'aide de l'amplification des gènes d'ARNr par des amorces spécifiques puis clonage et séquençage Sanger (Figures 4 et 5). Pour l'article 2, nous avons construit une première fois les OTU à partir des *reads* issus d'un premier *run* de séquençage. Par la suite, les OTU ont été reconstruites à partir de l'ensemble des séquences obtenues pour l'ensemble des 2 ans de suivis (Cf. Matériel et Méthodes général). Le premier jeu d'OTU (Cf. article chapitre 2), comprenait une OTU proche du groupe HAP-1, groupe d'haptophytes basaux uniquement détecté en eau douce (Šlapeta *et al.* 2005; Shalchian-Tabrizi *et al.* 2011). Les séquences

constituant cette OTU ont dû être intégrées à une autre OTU lors de la seconde construction d'OTU, et cette OTU n'existe plus.

En revanche, nous avons de nouveau pu détecter des séquences affiliées au nouveau groupe d'isochrysidales de l'étang des Vallées (clade EV; Figure 4). Les 3 OTU correspondantes (OTU 3305, 3828 et 3954) ont pu être amplifiées dans des échantillons de ce même étang uniquement, confirmant la préférence de ce groupe pour l'étang par rapport à d'autres écosystèmes géographiquement proches. Toutes méthodes confondues, ce groupe EV aura été détecté en avril 2009 dans le lac d'Annecy (OTU Prym\_1) et en février 2010 (OTU Prym\_1,2,3,4,19), mars 2011 (OTU Prym\_1), février et mars 2012 (OTU\_3305,3320,3954) et mars 2013 (OTU\_3305) dans l'étang des Vallées. Il apparaît donc comme un groupe récurrent de fin d'hiver dans cet écosystème. On notera cependant que seule l'OTU 3305, proche de l'OTU Prym\_19 et de séquences présentes dans les bases de données (100% d'identité sur 99,8% de couverture avec GU067829, séquence issue d'un lac luxembourgeois, notamment) est détectée au cours des 2 ans d'échantillonnage (Figure 6). Les OTU appartenant à la branche divergente ne sont retrouvées qu'en 2012. Cette branche présente une grande diversité, mais OTU 3954 est à 100% identique à EV3-Pry1-C11, séquence représentative de Prym\_4. Les nouvelles données moléculaires obtenues au cours du suivi bisannuel confirment donc le grand intérêt de ce groupe EV, et notamment de la branche très divergente



**Figure 6: Dynamique temporelle des 3 OTU affiliées au Groupe EV dans l'étang des Vallées.** Les triangles pleins correspondent aux échantillons de la fraction de taille cellulaire 0,2-5 µm, les triangles vides à la fraction 5-30 µm. L'identité des premiers hits par BLAST de la séquence représentative de chaque OTU, leur affiliation taxonomique et le pourcentage d'identité avec les séquences des OTU sont indiqués. Les OTU 3320 et 3954 correspondent à la branche divergente du groupe, dont les séquences n'ont pas de séquences proches dans les bases de données.

Enfin, l'étude des haptophytes, par clonage et séquençage Sanger d'amplicons d'ADNr 18S dans des échantillons prélevés ponctuellement, suggérait une faible abondance de ce groupe en eau douce, observation confortée par quelques analyses dans des lacs à l'aide d'amorces générales d'eucaryotes (Cf article). Nous avons envisagé que les organismes détectés puissent être rares au moment de l'échantillonnage mais devenir plus abondants à d'autres époques de l'année. Au cours des deux ans de suivi, les OTU d'haptophytes tous groupes confondus représentent toujours une faible proportion des séquences des bibliothèques dans lesquelles on les détecte (fréquence maximale : 0.30% dans un duplicat issu de l'Etang des Vallées en mars 2012, fréquence moyenne par échantillon : 0.08% des reads). De plus, elles ne sont détectées que dans 21 pools de séquences, correspondant à des échantillons de Saint Robert (3 dates), de l'Etang des Vallées (8 dates parmi les échantillons de cellules de la fraction de taille 0,2-5  $\mu\text{m}$  et 3 dates dont un échantillon en duplicats pour la fraction 5-30  $\mu\text{m}$ ), du ru Sainte Anne (4 dates) et de La Claye (2 dates), à différentes saisons. Les haptophytes n'ont jamais été détectés dans la mare Gabard durant le suivi bien que des pavlovophytes y aient été détectés en mars 2011 par clonage (Cf. article). Ainsi, ce groupe semble ne pas devenir abondant, à aucune saison, dans les écosystèmes étudiés. Toutefois, une étude à plus long terme pourrait éventuellement montrer des augmentations occasionnelles de leur abondance, suite à des événements ponctuels.





## Chapitre 2 : Exploration de la diversité des petits eucaryotes dans cinq écosystèmes d'eau douce peu profonds aux caractéristiques contrastées.

---



"Plus la tanière est petite, plus le blaireau y a chaud."  
Proverbe français



## I. Introduction : résumé du contexte et objectifs

L'étude précédente (chapitre 1) a confirmé la présence de groupes d'haptophytes encore à décrire dans les petits écosystèmes d'eau douce peu profonds, suggérant que ces systèmes pourraient héberger une part encore méconnue de la diversité des eucaryotes unicellulaires. Cette idée était confortée par la très forte sous-représentation des écosystèmes d'eau douce parmi les études de la diversité des eucaryotes en milieu aquatique, et des systèmes de petite taille (mares et ruisseaux) parmi les études en eau douce. Nous avons donc entrepris de partir à la découverte de cette diversité, ainsi que de la dynamique et éventuellement de l'écologie des communautés de petits eucaryotes, dans quatre mares et un ruisseau.

Les petits eucaryotes des mares et ruisseaux avaient jusqu'ici été très peu étudiés, notamment à l'aide de méthodes moléculaires. Avant d'approcher le fonctionnement des communautés de microorganismes eucaryotes, il nous est paru important de commencer par brosser un tableau général de leur diversité dans ces systèmes, à un instant "t". Nous avons donc examiné la diversité des petits eucaryotes dans des échantillons d'eau de surface prélevés en avril 2012, dans quatre mares et un ru (Cf. Matériel et méthodes général), par amplification à l'aide d'amorces générales d'eucaryotes et pyroséquençage 454 de fragments d'ADNr 18S. Nous avons ciblé les fractions de taille cellulaire 0,2-5  $\mu\text{m}$  pour l'ensemble des 5 écosystèmes, et 5-30  $\mu\text{m}$  pour l'étang des Vallées.

Cette première étape avait plusieurs objectifs principaux :

- Décrire de manière générale la diversité des eucaryotes microbiens dans les petits écosystèmes d'eau douce peu profonds, la comparer à celle observée dans les grands lacs et en milieu marin.
- Comparer la diversité observée entre écosystèmes géographiquement proches mais présentant des caractéristiques différentes (e.g. quantités de nutriments, pH, oxygénation). En effet, les petits écosystèmes d'eau douce peu profonds étant très divers, découvrir des communautés d'eucaryotes différentes dans différents systèmes même très proches laisserait supposer une très large diversité sur l'ensemble des mares et rus. D'autre part, comparer des écosystèmes différents nous permettrait également de tester l'influence relative de la distance et des conditions environnementales dans la répartition des organismes dans les mares et ruisseaux.
- Tester la fiabilité de notre méthode de détection de la diversité. Pour ce faire, tous les échantillons ont été réalisés et analysés indépendamment en duplicats, et les assemblages microbiens détectés dans chacun des deux duplicats d'une paire ont été comparés.

**II. Manuscrit de l'article 2 : "Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems"**

**Remarque :** En fin de manuscrit, un petit tableau récapitule les caractéristiques principales des différents échantillons (nom de l'échantillon, date et saison d'échantillonnage, présence de réplicats, nature eau / sédiment de l'échantillon). Il peut être déplié pour rester en vis-à-vis de toutes les pages du manuscrit.

*Article accepté par Environmental Microbiology*

**Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems**

**Marianne Simon, Ludwig Jardillier, Philippe Deschamps, David Moreira, Gwendal Restoux, Paola Bertolino, Purificación López-García\***

Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud, 91405 Orsay, France

\*For correspondence. E-mail: [puri.lopez@u-psud.fr](mailto:puri.lopez@u-psud.fr); Tel. +33 169 157 608; Fax +33 169 154 697

**Running title:** Protist diversity in shallow freshwater systems

**Keywords:**

Biodiversity, 18S rRNA, protist, plankton, freshwater, cryptophyte, MAST

## A- Summary

Although inland water bodies are more heterogeneous and sensitive to environmental variation than oceans, the diversity of small protists in these ecosystems is much less well-known. Some molecular surveys of lakes exist, but little information is available from smaller, shallower and often ephemeral freshwater systems, despite their global distribution and ecological importance. We carried out a comparative study based on massive pyrosequencing of amplified 18S rRNA gene fragments of protists in the 0.2-5  $\mu$ m-size range in one brook and four shallow ponds located in the Natural Regional Park of the Chevreuse Valley, France. Our study revealed a wide diversity of small protists, with 812 stringently defined operational taxonomic units (OTUs) belonging to the recognized eukaryotic supergroups (SAR –Stramenopiles, Alveolata, Rhizaria–, Archaeplastida, Excavata, Amoebozoa, Opisthokonta) and to groups of unresolved phylogenetic position (Cryptophyta, Haptophyta, Centrohelida, Katablepharida, Telonemida, Apusozoa). Some OTUs represented deep-branching lineages (Cryptomycota, Aphelida, Colpodellida, Tremulida, clade-10 Cercozoa, HAP-1 Haptophyta). We identified several lineages previously thought to be marine including, in addition to MAST-2 and MAST-12, already detected in freshwater, MAST-3 and possibly MAST-6. Protist community structures were different in the five ecosystems. These differences did not correlate with geographical distances, but seemed to be influenced by environmental parameters.



## B- Introduction

Aquatic ecosystems occupy most of the Earth surface. Oceans alone cover around 71% of that surface (Costanza 1999). Lakes, ponds and reservoirs cover more than 3% (nearly 4.5 million km<sup>2</sup>) of non-oceanic regions (Downing *et al.* 2006). Although comparatively smaller, these freshwater systems offer a large array and diversity of ecological niches including various trophic levels, light accessibility, temperature and oxygen concentrations. This is especially true for small and shallow lentic inland ecosystems (sizes from 0.001 to 0.1 km<sup>2</sup>), which are widespread, varied and numerous, corresponding to more than 99% of the total number of lakes on earth (Downing *et al.* 2006). Microbial communities dominate aquatic ecosystems and their activity has profound impact at global scales, being largely implicated in carbon fixation (Li 1994; Jardillier *et al.* 2010) and climate regulation (Simó 2001). Within these communities, microbial eukaryotes play key roles in nutrient cycling acting as photosynthesizers, heterotrophs (predators, parasites) or mixotrophes (Caron 1994; Zubkov & Tarran 2008; Jardillier *et al.* 2010; Massana 2011).

Small eukaryotes (<20 µm) have been known to constitute a non-negligible part of aquatic microbial communities in both freshwater and marine systems for a long time (Johnson & Sieburth 1982; Corpe & Jensen 1992). Based on their cell size, small protists were initially classified in nanoeukaryotes (cells between 2 and 20 µm in diameter) and picoeukaryotes (cells ≤ 2 µm). In the last two centuries, many small eukaryotic species have been described, including phototrophs such as prasinophytes and other chlorophytes (Knight-Jones *et al.*, 1951; Stockner, 1988; Guillou *et al.*, 1999), which suggested a potentially significant role in primary production (Johnson & Sieburth 1982), but also heterotrophs (Fenchel, 1982; Patterson and Larsen, 1991). Indeed, the ecological importance of heterotrophic nanoflagellates as predators has long been acknowledged (Fenchel 1982; Wright & Coffin 1984). However, being too small to show easily-identifiable, unambiguous morphological differences, their true diversity remained largely inaccessible and their taxonomy poorly studied or oversimplified (many of these tiny protists were simply classed as *incertae sedis*).

In the past 15 years, the use of molecular methods based on 18S rRNA gene analysis has largely overcome that problem, allowing studying the phylogenetic diversity and distribution of small protists at unprecedented scales. Most such studies largely explored oceanic systems, including surface waters (Diez *et al.* 2001; Moon-van-der-Staay *et al.* 2001), the deep-sea (López-García *et al.* 2001), and coastal regions (Romari & Vaultot 2004; Massana *et al.* 2004a; Cheung *et al.* 2010). They revealed a huge protist diversity encompassing members of

all recognized eukaryotic super-groups (SAR—Stramenopiles, Alveolata, Rhizaria—, Archaeplastida, Excavata, Amoebozoa and Opisthokonta as well as lineages of uncertain position in the eukaryotic tree, such as haptophytes, cryptophytes or picozoa (López-García & Moreira 2008; Massana 2011; Seenivasan *et al.* 2013; Moreira & López-García 2014). Molecular surveys also allowed the discovery of lineages previously unknown in spite of their abundance, such as new clades affiliated to alveolates (e.g. Groups I and II), which turned out to be members of the classical Syndiniales (Groissillier *et al.* 2006; Harada *et al.* 2007; Guillou *et al.* 2008) and stramenopiles (the MAST clades, for MArine STRamenopiles) (López-García *et al.* 2001; Moon-van der Staay *et al.* 2001; Diez *et al.* 2001; Stoeck *et al.* 2003; Massana *et al.* 2004). Thought for a long time to be exclusively marine, the MAST groups remain poorly known. In a few cases a correspondence has been found between particular organisms and their environmental sequences (e.g. the colonial protist *Solenicola setigera* and the MAST-3; Gómez *et al.* 2011). However, many other MAST lineages have not yet been linked to any cultured or described organisms (Massana *et al.* 2013), even though fluorescent in situ hybridization (FISH) labeling has provided some hints on their morphology, life style and ecology (Massana *et al.* 2002, 2006).

The molecular exploration of very small protists in freshwater began slightly later than in the oceans and also revealed a large diversity (Lefranc *et al.* 2005; Richards *et al.* 2005; Šlapeta *et al.* 2005). In lakes, alveolates (especially ciliates and Perkinsozoa, by contrast to the Syndiniales-Duboscquellids-dinoflagellate dominance in marine systems), stramenopiles, cryptophytes and fungi were found to be abundant, but protists belonging to the Archaeplastida, Rhizaria, and Cercozoa or groups of uncertain affiliation were also diverse (Lefranc *et al.* 2005; Šlapeta *et al.* 2005; Lepère *et al.* 2008; Zhao *et al.* 2011; Mangot *et al.* 2012). In some cases, a significant part of the detected diversity was composed only of environmental sequences without known close relatives in databases (Lefèvre *et al.* 2008; Mangot *et al.* 2012; Triadó-Margarit & Casamayor 2012). Some groups were observed in both oceans and freshwater systems while others seem, according to current knowledge, to be specific to marine (e.g. Syndiniales; Guillou *et al.* 2008) or freshwater (e.g. HAP-1 lineage of haptophytes; Šlapeta *et al.* 2005; Shalchian-Tabrizi *et al.* 2011) environments. Even within lineages present both in marine and freshwater systems, 18S rRNA sequences obtained from freshwater samples often form phylogenetic clades distinct from those of oceanic sequences, suggesting a limited number of, mostly ancient, freshwater colonization events followed by radiations (Logares *et al.* 2009). That scarcity of marine-freshwater transitions has been observed for many small eukaryote groups such as the perkinsids (Bråte *et al.* 2010b), haptophytes (Shalchian-Tabrizi *et al.* 2011; Simon *et al.* 2013) or MAST lineages (Massana *et al.*

*al.* 2013), and could be explained by the difficulty of crossing the salinity barrier (Logares *et al.* 2009). However, freshwater aquatic systems remain largely understudied and massive high-throughput sequencing techniques have been applied to very few among them. Therefore, it might be possible that failure to detect some of these lineages is due to undersampling. Furthermore, the vast majority of protist molecular diversity studies in freshwater have been conducted in lakes (eg. Richards 2005, Lepère *et al.* 2006, 2008). However, smaller systems like ponds or brooks have been overlooked, despite their abundance, distribution and ecological importance. Yet, previous analyses suggest that they host several lineages undetected in marine environments or lakes (Šlapeta *et al.* 2005; Simon *et al.* 2013). In addition, because they are shallow and small, this kind of freshwater systems may display very different local physico-chemical conditions and, hence, they constitute ideal models to test whether local environmental selection is more influential than geographic distance in determining community composition, which remains controversial (Green *et al.* 2004; Martiny *et al.* 2006). A recent study on lakes suggest that geographic distance might be important to explain protist biogeography (Lepère *et al.* 2013), but lakes have much more buffering capacity than small shallow systems.

In this work, we have investigated the diversity of small eukaryotes (essentially the 0.2-5 µm size-fraction) in a set of shallow freshwater environments by amplification of 18S rRNA gene fragments and direct high-throughput 454-pyrosequencing. We selected one brook and four ponds located in the same geographic area (the Natural Regional Park of the Chevreuse Valley, France) but differing in size, depth and physico-chemical conditions. The objectives of our work were threefold: i) describing protist diversity at unprecedented depth in this kind of habitats, ii) checking whether increasing sequence depth leads to the discovery in freshwater systems of eukaryotes previously thought to be exclusively marine and iii) testing which parameters (geographical distance vs. physico-chemical characteristics) determine eukaryotic microbial communities.

## C- Results

To study protist diversity in five small and shallow freshwater ecosystems, we selected one brook and four shallow ponds in the Natural Regional Park characterized by different physico-chemical parameters and trophic status (Table 1; Fig. S1).

The Mare Gabard is a small pond located in the middle of the forest (Fig. S1). It smelled strongly of H<sub>2</sub>S when sediments were stirred and had high phosphate concentrations

(0.15 mg.l<sup>-1</sup>) but the lowest pH, conductivity and TDS (total dissolved solids) values. Saint Robert pond is the most anthropized system selected due to its location in a hamlet (Fig. S1). It contained very high ammonia (1.22 mg.l<sup>-1</sup>) and, to a lesser extent, also nitrite concentrations (Table 1, Fig. 1), likely influenced by the permanent presence of a sizeable duck population. La Claye pond is part of a complex of several ponds on ancient peat bog substrate, with a dense population of eagle ferns in the surroundings. Lying on acidic and organic-rich soils, it was characterized by high concentrations of dissolved organic carbon (36.3 mg.l<sup>-1</sup>).

*Table 1: Characteristics of studied freshwater systems, and diversity and richness estimates for the different samples. Replicate samples are indicated by a small case letter after the sample name. TDS, total dissolved solids; DOC, dissolved organic carbon.*

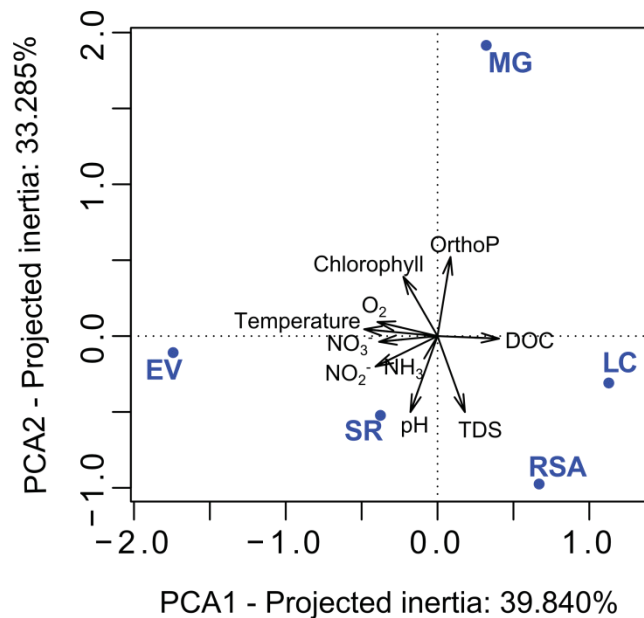
Sampled ecosystem	Gabard		Saint Robert		Etang des Vallées				Sainte Anne		La Claye	
Ecosystem type	Forest pond		Village pond		Large pond				Forest brook		Pond on peaty soil	
GPS coordinates	48°39'15.83"N 1°55'20.26"E		48°39'54.82"N 1°56'45.28"E		48°41'23,0"N 001°54'59,2"E				48°36'45.91"N 1°58'16.61"E		48°36'31.72"N 1°56'17.33"E	
Approximate surface (m <sup>2</sup> )	850 (210 x 75 m)		495 (20 x 28 m)		12 880 (210 x 75 m)				1 m width at sampling point		265 (24 x 10 m)	
Approximate depth (cm)	50		50		150				20		60	
Sampling Date	April 5, 2012		April 5, 2012		March 30, 2012				April 6, 2012		April 6, 2012	
Water temperature (°C)	9.0		10.7		13.1				7.1		7.8	
pH	6.6		7.2		7.31				7.36		7	
Oxygen (%)	78.9		58.9		116.8				72.8		61.2	
Conductivity (µS/cm)	81.4		531		288				746		520	
TDS (mg/l)	81		531		288				746		520	
Chlorophyll (µg /l)	74.9		61.9		44.7				2.6		10.7	
NO <sub>3</sub> <sup>-</sup> (mg/l)	1.41		1.17		5.08				1.45		2.02	
NO <sub>2</sub> <sup>-</sup> (mg/l) <sup>2</sup>	~ 0		0.054		0.047				0.010		~ 0	
NH <sub>3</sub> (mg/l)	0.02		1.22		0.02				0.03		0.04	
DOC (mg/l)	19.0		15.7		9.8				17.7		36.3	
PO <sub>4</sub> <sup>3-</sup> (mg/l <sup>-1</sup> )	0.15		0.03		0.03				0.03		0.03	
Size fraction studied	0.2 - 5 µm		0.2 - 5 µm		5 - 30 µm	0.2 - 5 µm			0.2 - 5 µm		0.2 - 5 µm	
Sample Name	MG25	MG25b	SR25	SR25b	EV34	EV34b	EV33	EV33c	RSA25	RSA25b	LC25	LC25b
Number of reads before filtering	15,020	5,110	85,803	6,724	28,418	14,168	22,314	7,938	37,766	7,140	21,283	14,215
Number of quality-filtered reads	10,616	4,197	42,034	4,506	10,982	6,906	17,670	3,947	19,652	4,243	11,191	10,605
Observed number of OTUs	76	55	15	37	147	132	177	87	427	198	66	68
Richness <sup>1</sup>	67.9	55.0	11.4	36.8	134.4	126.9	148.5	87.0	313.7	197.8	57.9	57.0
(Standard error)	(2.37)	(0.15)	(1.06)	(0.44)	(2.94)	(2.07)	(3.99)	(0.00)	(7.52)	(0.44)	(2.34)	(2.67)
Diversity (Simpson Index)	0.86	0.80	0.75	0.68	0.96	0.95	0.95	0.95	0.92	0.91	0.50	0.38
Evenness	0.58	0.56	0.61	0.45	0.75	0.72	0.69	0.78	0.64	0.7	0.34	0.27

<sup>1</sup> Expected number of OTUs in random subsamples of the size of the smallest sequence library (3,947 reads in EV33c)

<sup>2</sup> Nitrite concentrations in Gabard and La Claye ponds were under the kit detection limit of 0.006 mg/l and were set as 0 in analysis.

## Chapitre 2

The Etang des Vallées is a nearly 1.5 m deep pond and, contrary to other sampled systems, was supersaturated in oxygen (116.8%; Table 1). It also had high temperature at the time of sampling (13.1°C), and high concentrations of nitrate (5.1 mg.l<sup>-1</sup>) as compared to the other systems. The Ru Sainte Anne is the brook and had the highest amounts of total dissolved solids (746 mg. l<sup>-1</sup>) possibly due to sediment input in a very shallow (~10 cm) system of running waters (Table 1). Principal component analysis (PCA) of the different physico-chemical parameters measured showed that, despite their close proximity (distances between these systems varied from 2 to 9 km; Fig. S1), the ponds and brook were clearly distinct from each other (Fig. 1). This made of this set of shallow aquatic ecosystems a good model to study whether geographic proximity is more influential than physico-chemical parameters in determining community composition.

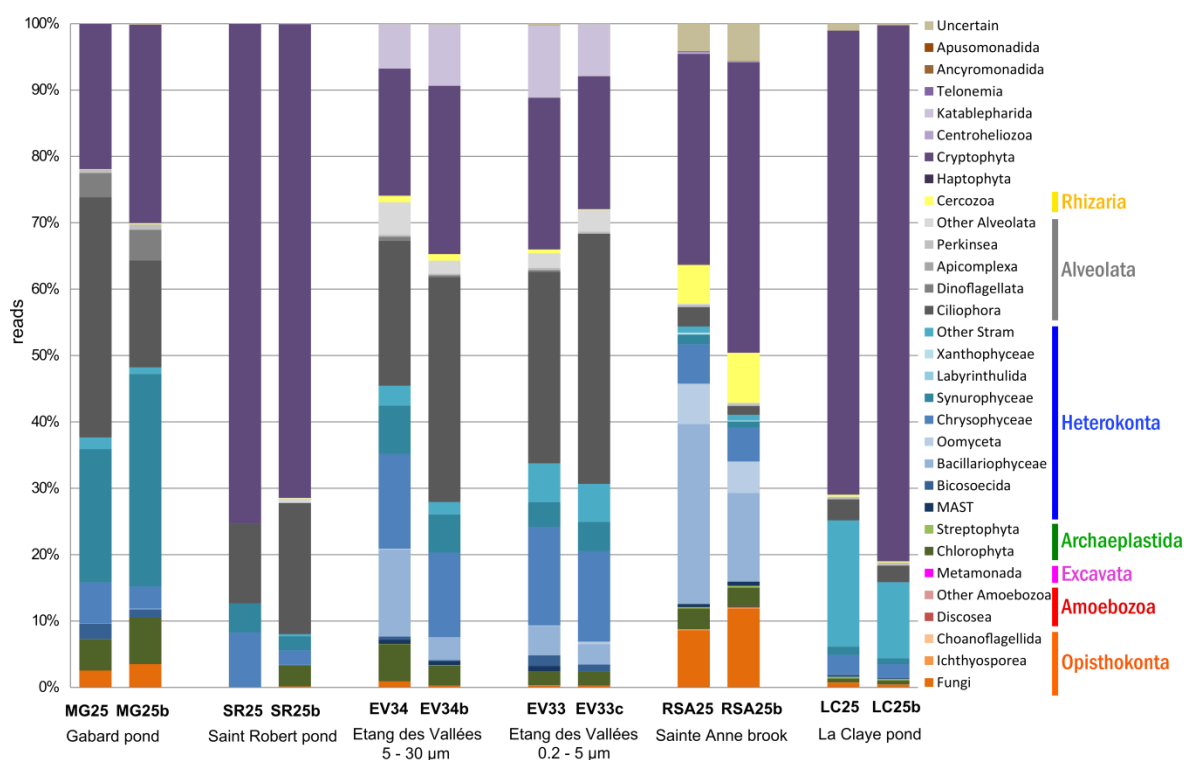


**Figure 1: Principal Component Analysis (PCA) plot of the measured physicochemical parameters.** Sampled ecosystems appear in blue and physico-chemical parameters in black. MG, Mare Gabard; EV, Etang des Vallées; LC, La Claye; RSA, Ru Sainte Anne; SR, Saint Robert. TDS, Total Dissolved Solutes; DOC, Dissolved Organic Carbon; OrthoP, orthophosphate.

### Overall protist community composition

The composition of protist communities was estimated for the five selected shallow freshwater systems based on 454-pyrosequenced 18S rDNA fragments amplified from DNA of plankton of the 0.2-5 µm cell diameter fraction. In addition, protist diversity in the size fraction 5-30 µm was studied for the largest sampled ecosystem, the Etang des Vallées. Replicate samples were included in all cases. We described protist diversity in these systems from a total of 146,549 quality-filtered reads using highly stringent criteria to avoid sequence artefacts and chimeras. The reads from the 12 different samples (the 0.2-5 µm cell fraction of 5 systems with replicas plus the two replicas of the 5-30 µm for the Etang des Vallées) were treated together in order to define operational taxonomic units (OTUs) that could be fully compared among samples. All those sequences grouped in 812 OTUs defined using a cut-off

value of 98% sequence identity. From these, 768 OTUs (128,661 reads) were detected at least in the 0.2-5  $\mu\text{m}$  size fraction, the remaining 44 OTUs corresponding to protists found only in the larger 5-30  $\mu\text{m}$  at the Etang de Vallées (Table 1). The different OTUs were then assigned to known taxonomic groups based on sequence similarity, which revealed a wide phylogenetic diversity of protists in general and of small eukaryotes in particular. Our stringently defined OTUs affiliated to the major recognized eukaryotic supergroups SAR (Stramenopiles-Alveolata- Rhizaria), Archaeplastida, Excavata, Amoebozoa and Opisthokonta (López-García & Moreira 2008; Adl *et al.* 2012) and to several lineages of unresolved phylogenetic position such as Cryptophyta, Haptophyta and Apusozoa (Fig. 2).



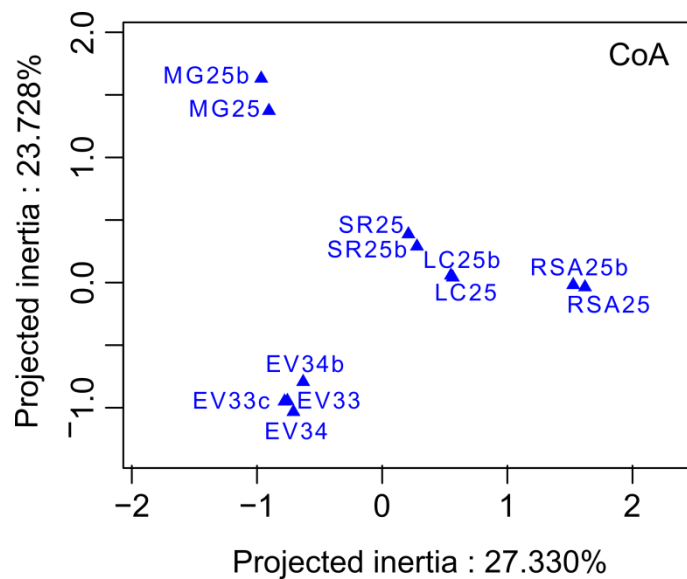
**Figure 2:** Histogram showing the relative proportion of 18S rRNA gene amplicon reads assigned to high-rank taxa in the five shallow ecosystems studied. Replicate samples are labeled as 'b' or 'c'. In all cases, the distribution corresponds to protists in the 0.2-5  $\mu\text{m}$  size range, except in the Etang des Vallées, where the 5-30  $\mu\text{m}$  size fraction was additionally analyzed.

To evaluate the reliability of the community composition determined for each ecosystem at this stage, we compared the pairwise Bray-Curtis distances between the two replicates for each ecosystem. These were very small (0.26 on average, min. 0.12, max. 0.36) as compared to distances between libraries from distinct ecosystems (0.84 on average, min. 0.44, max. 0.99). Replicates grouped together in both Non-metric MultiDimensional Scaling (NMDS;



Fig. S2) and Correspondence (CoA) analyses (Fig. 3), revealing their high similarity in OTU composition.

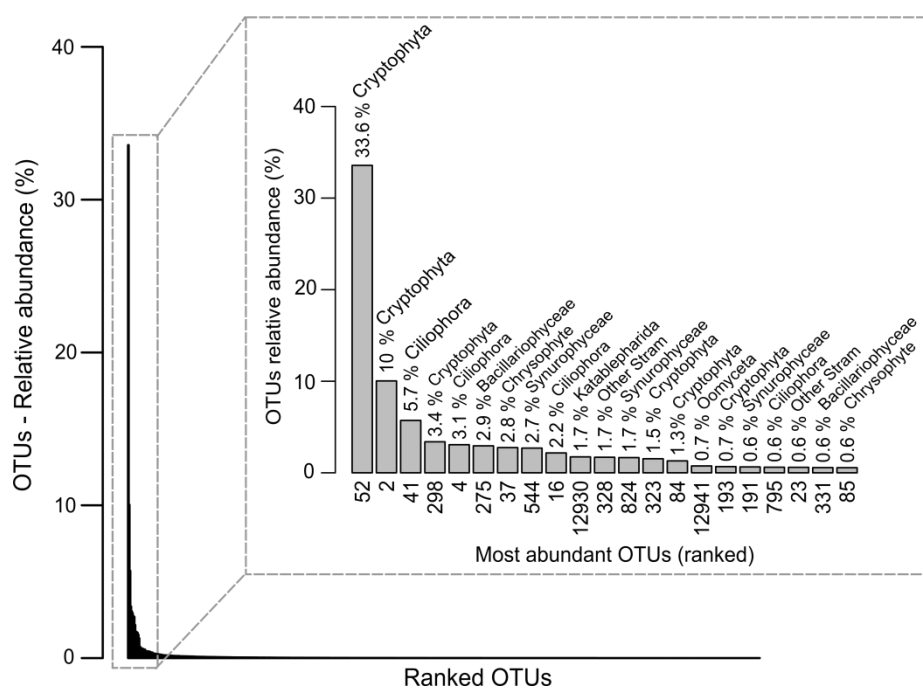
**Figure 3: Correspondence Analysis (CoA) plot showing protist community composition similarities and differences among the five ecosystems studied.** MG25 and MG25b, Mare Gabard; LC25 and LC25b, La Claye; SR25 and SR25b, Saint Robert; RSA25 and RSA25b, Ru Saint Anne; EV33 and EV33c, Etang des Vallées (0.2-5  $\mu\text{m}$  size samples). EV34 and 34b, Etang des Vallées (5-30  $\mu\text{m}$  size samples).



The composition of protist communities from the small size fraction samples (0.2-5  $\mu\text{m}$ ) differed greatly between ecosystems (Fig. 2). First, richness and diversity indexes were highly variable (Table 1). Samples from Ru Sainte Anne and Etang des Vallées were the richest and the most diverse, whereas La Claye was highly dominated by few abundant OTUs (evenness = 0.34 – 0.27 in replicates). La Claye and Saint Robert appeared relatively close in NMDS plots (Fig. S2) and clustered close together in CoA plots (Fig. 3). These two systems displayed relatively similar physico-chemical parameters, notably oxygen content and conductivity, and low diversity as compared to the other ecosystems (Table 1). The similarity between La Claye and St Robert community composition was likely influenced by the fact that they were largely dominated by the same cryptophyte OTU (Fig. 2; see below).

Second, the relative abundance of taxonomic groups varied between ecosystems, even though OTUs affiliated to Cryptophyta, Stramenopiles (or Heterokonta) and Alveolata accounted for the majority of sequences in all libraries (Figs. 2 and 4). Cryptophyte sequences were remarkably abundant, representing up to 49 and 53% of the total number of reads and reads of the 0.2-5  $\mu\text{m}$  size fraction respectively. Cryptophytes were the dominant group in Saint Robert and La Claye (around 75% of reads in those samples). They were also the dominant group in the brook Sainte Anne (between 32-44% of reads), and were the second most abundant group in the two other systems (Fig. 2, Table S1). Most cryptophyte reads belonged to a unique OTU affiliated to *Cryptomonas curvata* (OTU\_52; Figs. 4 and S3). Stramenopiles constituted the second most abundant group in our systems and represented 25% of all reads and nearly 24% of the reads from the 0.2 -5  $\mu\text{m}$  size fraction samples (Fig. 2,

Table S1). Within this supergroup, Chrysophyceae were abundant in all samples. However, the most abundant stramenopile groups in Mare Gabard and Ru Sainte Anne were, respectively, Synurophyceae and Bacillariophyceae, two lineages producing silica skeletons or scales. Oomycetes were also relatively abundant in the brook Sainte Anne. Alveolates constituted the third most abundant supergroup in our study. It was the dominant group in Etang des Vallées (32-41% of reads in replicates of the 0.2-5  $\mu\text{m}$  size fraction, and similar values for the larger size fraction analyzed) and one of the dominant groups in Mare Gabard (22-40% of reads in duplicate samples). It was also a major component in Saint Robert pond (12-20%). However, they represented only a small proportion of the taxa detected in the brook Sainte Anne and La Claye pond (less than 5 % of reads; Fig. 2, Table S1). Among alveolates, ciliates were the most represented in all samples; though OTUs affiliated to dinoflagellates and other alveolates had occasionally significant proportions.



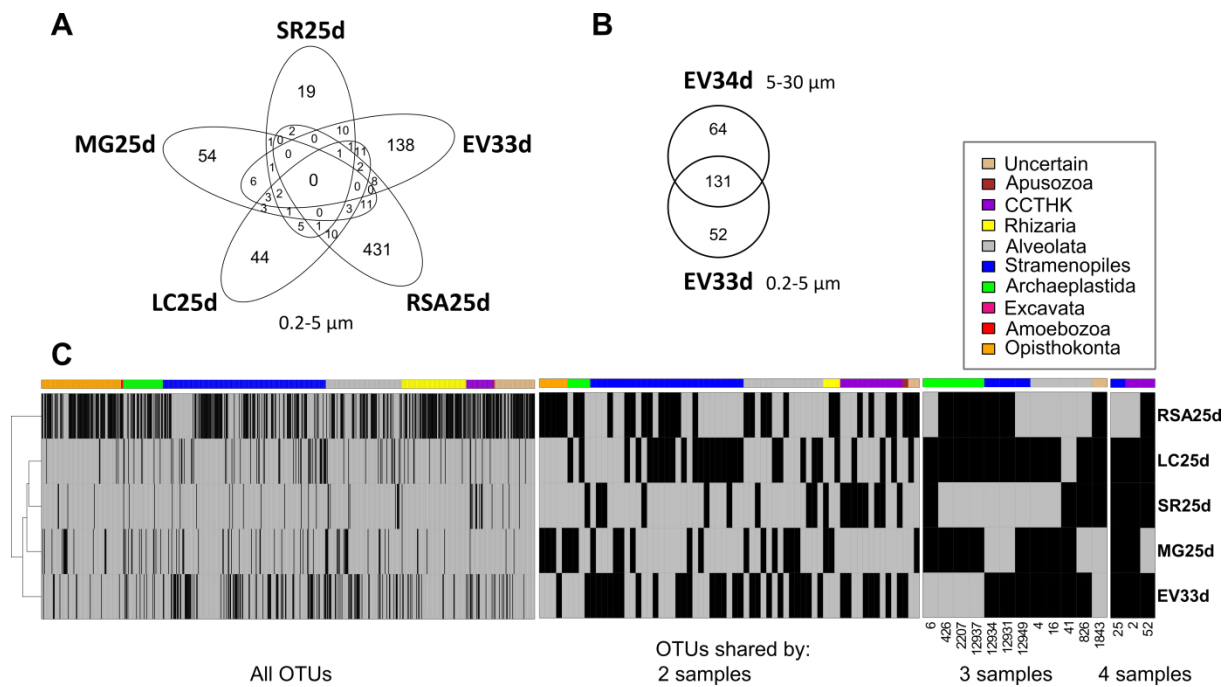
**Figure 4:** Rank abundance curve for the total 768 OTUs detected collectively in the 0.2-5  $\mu\text{m}$  size fraction of the five shallow freshwater systems studied. Sequence data from all samples were pooled to define OTUs with high stringency. The relative abundance of protist OTUs representing more than 0.5% of the total number of reads are shown in the inset. The identity number of the respective OTUs and their taxonomic affiliation are shown below and above the histogram bars.

Along with these three abundant supergroups, members of the Opisthokonta, Amoebozoa, Excavata, Archaeplastida, Rhizaria and of groups of uncertain position in the eukaryotic tree were also detected. Since we purposefully used general eukaryotic primers biasing against Metazoa, opisthokonts were essentially represented by fungi, though

choanoflagellates and ichthyosporeans were also detected. The highest abundance of fungi was observed in Mare Gabard and Ru Sainte Anne, but they generally represented less than 10% of reads. Rhizarian OTUs were retrieved in all ecosystems and were only represented by cercozoans. They were not abundant except for the Ru Saint Anne, where they reached 6-8% of the reads. Katablepharid sequences represented as much as 8-11% of reads from the 0.2-5  $\mu\text{m}$  fraction in the Etang des Vallées but were less abundant elsewhere (0.03 % on average in each small size fraction sample). Archaeplastida (streptophytes and, mainly, chlorophytes) were detected in all ecosystems. Haptophytes were also identified, though in low proportions, and only in the Etang des Vallées and the Ru Sainte Anne. OTUs affiliated to Excavata, Labyrinthulida, Xanthophyceae, Apicomplexa, Centrohelioczoa, Telonemida, Amoebozoa and Apusomonadida (Apusozoa) were detected only in the highly diverse Sainte Anne brook (Fig. 2, Table 1). 3 additional apusozoan OTUs were detected in Sainte Anne brook and the Etang des Vallées.

Despite some similarity in the distribution of large phylogenetic groups, with cryptophytes, stramenopiles and alveolates dominating the different shallow water systems, protist communities were very different at the phylotype scale. Indeed, no OTU was shared by all the systems and the vast majority of the 768 OTUs detected in the small size fraction was specific to the libraries of a single ecosystem (Fig. 5A). 67 and 12 OTUs were shared by 2 and 3 different ecosystems, respectively (Fig. 5C). Only three OTUs were shared by four systems (Fig. 5C), two of which affiliated to cryptophytes and corresponded to the most abundant OTUs (Fig. 4). Remarkably, OTU\_52, affiliated to *Cryptomonas curvata*, was particularly abundant in St Robert and La Claye samples where it represented as much as 43-52% and 69-78% of reads respectively. However, OTU\_52 dropped to 4-7% of reads in the 0.2-5  $\mu\text{m}$  fraction of the Etang des Vallées and was not detected in Mare Gabard. OTU\_25, affiliated to stramenopiles, was also shared by four systems, but represented only 0.13% of all reads in 0.2-5  $\mu\text{m}$  fractions.

The composition of protist communities within the two different size fractions (0.2-5  $\mu\text{m}$  and 5-30  $\mu\text{m}$ ) of the Etang des Vallées was very similar, as revealed by the NMDS and CoA analyses (Figs. 3 and S2) as well as by low Bray-Curtis distances (0.35 on average, min: 0.29, max: 0.43). Moreover, the community structure was similar in both size fractions from the Etang des Vallées, which displayed the highest diversity and evenness recorded of all samples (Table 1). The taxonomic composition was also similar at high-rank taxa level, although diatoms and green algae were in slightly higher proportions in the largest size fraction (Fig. 2). At a finer level of resolution, more OTUs were shared by both size fractions (131 OTUs) than specific to each of them (Fig. 5B).



**Figure 5: Distribution of protist OTUs in the five shallow freshwater systems studied.** Sequence data from replicates were pooled. A, five-set Venn diagram showing the number of specific and shared OTUs in the different freshwater systems (0.2-5  $\mu\text{m}$  fraction size). B, Venn diagram showing OTUs shared by the two fraction sizes analyzed in Etang des Vallées. C, clustering analysis of the five ecosystems based on the presence of shared OTUs, as shown by the heatmap. Only 0.2-5  $\mu\text{m}$  size fractions are considered. Heatmaps show all OTUs and OTUs shared by 2, 3 or 4 ecosystems. No OTU is shared by all the 5 ecosystems. Each row represents an ecosystem and each vertical bar an OTU. Black: OTU present, Grey: OTU absent. The colored sidebar indicates the taxonomic affiliation of the OTU represented by the bar below. Color codes representing different phylogenetic affiliation are indicated in the box. CCTHK: Cryptophyta, Centroheliozoa, Telonemia, Haptophyta and Katablepharida. For presentation reasons, the width of OTU bars is not the same in all the heatmaps. RSA25d, Ru Sainte Anne; LC25d, La Claye; SR25d, Saint Robert; MG25d, Mare Gabard; EV33d, EV34d, Etang des Vallées.

### Phylogenetic diversity

In order to get more detailed information on the different OTUs identified, we carried out phylogenetic analyses with partial 18S rRNA gene sequences representative of the different OTUs. Although cryptophytes were the most abundant group in our samples, they were not very diverse. Only 27 OTUs were affiliated to that group (3.3% of all OTUs). Most of these OTUs were closely related to existing sequences, affiliating to photosynthetic genera, especially *Cryptomonas* (containing the highly overrepresented OTU\_52 corresponding to *C. curvata* mentioned above) and *Chroomonas*, but also to the phagocytic plastid-lacking *Goniomonas* (Fig. S2). However, several OTUs appeared to be quite divergent; for instance representative sequences of OTUs 838 and 1812 shared no more than 93% of identity with their first BLAST hit in the NCBI database, the sequences HM135076 from freshwater (Luo

*et al.* 2011) and AM901364 from the cultured *Cryptomonas commutata* strain M1975 respectively (Fig. S3).

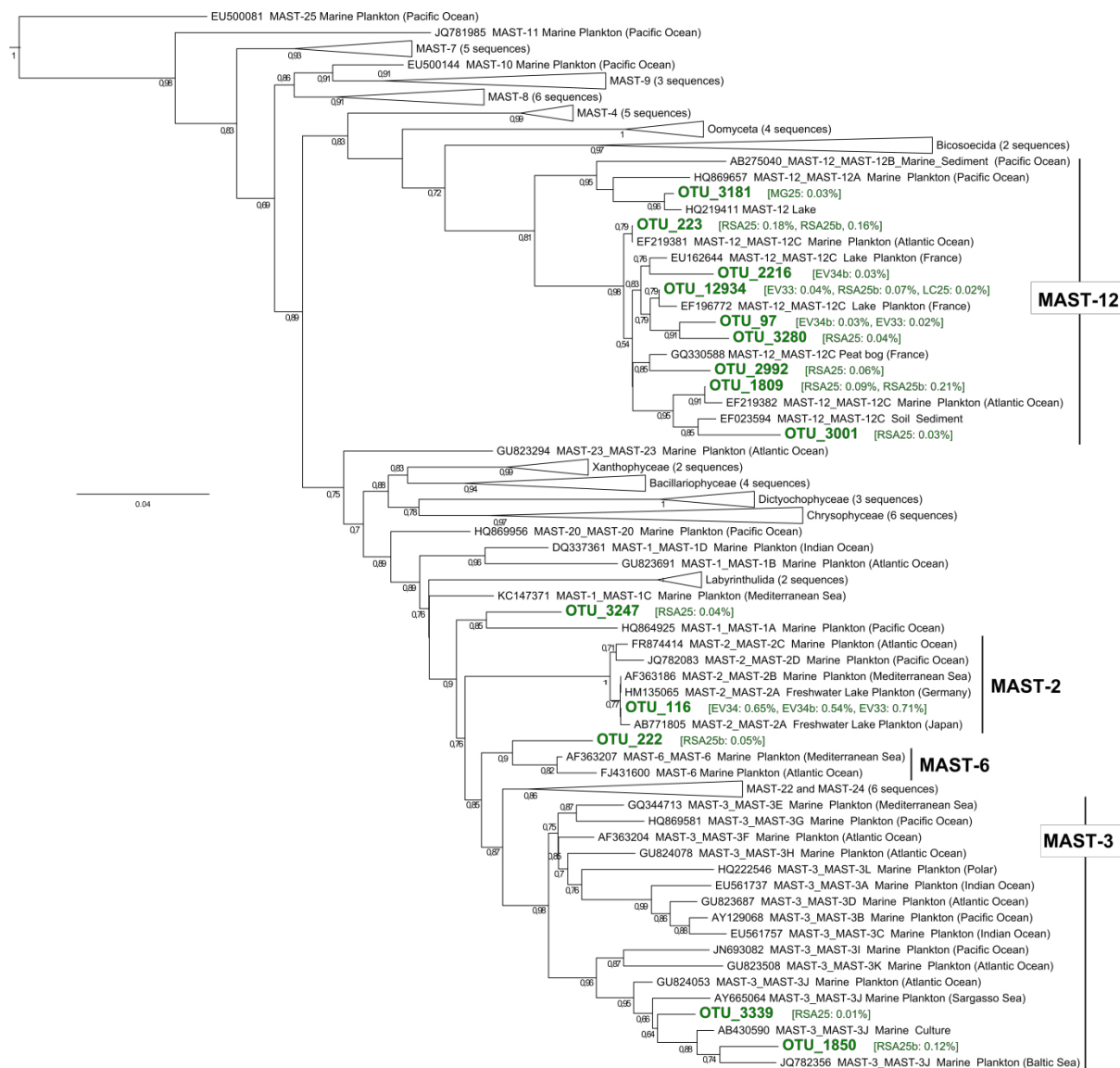
A total of 268 OTUs (33% of all OTUs) were affiliated with stramenopiles, the second most abundant group in our samples. Nearly half of them were related to Chrysophyta-Synurophyta (125 OTUs) while the remaining ones affiliated to Eustigmatophyta, Dictyochophyta, Bicosoecida, Oomycetes, Bacillariophyta, Xanthophyta and Labyrinthulida (Fig. S4). Surprisingly, 14 OTUs seemed to belong to various MAST groups (Fig. 6), originally thought to be exclusively marine. Nine OTUs affiliated to the group MAST-12, being related to sequences previously detected in a wide variety of ecosystems, such as a suboxic Norwegian estuary (Kolodziej & Stoeck 2007), freshwater lakes (Lefèvre *et al.* 2008; Monchy *et al.* 2011) and a peat bog (Lara *et al.* 2011). OTU\_116 was closely related (99% identity) to MAST-2 sequences detected in a freshwater lake (Luo *et al.* 2011) and the Mediterranean (Diez *et al.* 2001). OTUs\_1850 and 3339 affiliated to MAST-3, which so far was known to contain only sequences from marine environments. Finally, OTU\_222 and OTU\_3247 had similarity by BLAST to MAST-6 sequences retrieved from the Mediterranean (AF363207, Diez *et al.*, 2001) and other MAST-6 sequences, with 94% and 91% identity, respectively. However, from the phylogenetic analysis, their affiliation to this clade is unclear and will require a more in-depth exploration of these groups.

Alveolates were also found to be diverse (Figs. S5 and S6), but the vast majority of OTUs were assigned to ciliates (88 out of 125 alveolate OTUs). Dinoflagellates were also represented in our samples, along with 17 OTUs related to sequences of putative freshwater (Amaral-Zettler *et al.* 2008; Monchy *et al.* 2011) or marine (Scheckenbach *et al.* 2010; Behnke *et al.* 2010) perkinsid parasites. Interestingly, several OTUs grouped with Apicomplexa and Colpodellida and clustered into 2 distinct groups (Fig. S5). Eight OTUs clustered with *Colpodella edax* (Leander *et al.* 2003) and many environmental sequences from freshwater lakes or ponds (Richards *et al.* 2005; Lefèvre *et al.* 2007; Oikonomou *et al.* 2012; Nakai *et al.* 2012). Three other OTUs clustered with sequences affiliated to apicomplexans (mostly *Cryptosporidium*) coming from more diverse environments, e.g. peat bog (Lara *et al.* 2011), marine sediment (Dawson & Pace 2002), humans (Yuan *et al.* 2012) or ostrich fecal samples (Martinez-Diaz, R, unpublished).

From the 131 OTUs affiliated with opisthokonts, the great majority (125 OTUs) related to fungi, mostly to chytrids but also to ascomycetes, basidiomycetes or to the basal Rozellida/Cryptomycota lineage (Fig. S7). In addition 11 OTUs were related to the aphelids, the sister group to fungi, which so far mostly contains highly divergent 18S rRNA gene



environmental sequences (Karpov *et al.* 2014). Several OTUs clustered with ichthyosporeans and choanoflagellates (Fig. S7).



**Figure 6: Approximate Maximum Likelihood (ML) phylogenetic tree of partial 18S rRNA gene stramenopile sequences showing the presence of several MAST clades in shallow freshwater systems. Non-MAST stramenopile lineages have been collapsed. A total of 392 unambiguously aligned positions were used to reconstruct the tree. Two alveolate sequences were used to root the tree. Representative sequences of OTUs from this work are shown in bold green. The name of samples where MAST OTUs were detected and the respective proportion of reads are shown within brackets. The scale bar represents the number of estimated substitutions per position for a unit branch length. RSA25/RSA25b, Ru Saint Anne; MG25/MG25b, Mare Gabard; EV33/ EV33b (0.2-5  $\mu$ m) and EV34/EV34b (5-30  $\mu$ m), Etang des Vallées.**

All rhizarian OTUs branched with cercozoan sequences. Even though cercozoan OTUs were not very abundant in our samples and represent only 1.2% of the total number of reads,



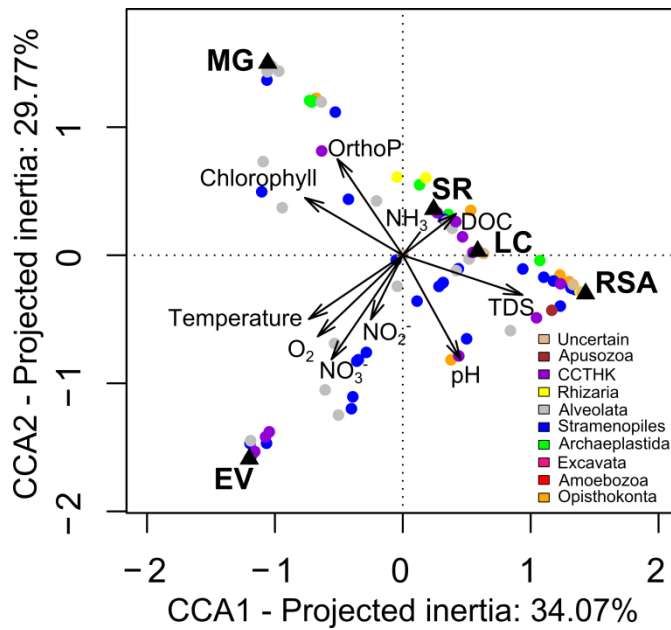
they were composed of numerous phylotypes (106 OTUs) that represented altogether 13% of all OTUs. The vast majority of these OTUs affiliated to the Filosa (Fig. S8). Several OTUs branched with the Endomyxa, although they shared only 93.6% identity on average with their first BLAST hits in GenBank (calculated on the 13 OTU representative sequences affiliated to Endomyxa, min: 89% between OTU\_12985 and AB526843, max: 99% between OTU\_1810 and EU910610). Furthermore, 5 OTUs clustered with sequences from the Novel Clade 10, and OTUs\_2162 and 608 shared 92% and 91% identity respectively with sequence EU567287 from Novel Clade 11 – Tremulida, two recently defined deep-branching cercozoan lineages (Bass *et al.* 2009; Howe *et al.* 2011).

65 OTUs (8% of all OTUs) belonged to Archaeplastida (Fig. S9). Five of them strongly affiliated to Embryophyta. Given the pre-filtration steps, these OTUs must correspond to pollen grains, free DNA or cells from dead leaves. Strikingly, they represented altogether only 0.04% of all reads, in spite of the dense vegetation around most sampled systems. Four additional OTUs most likely corresponded to unicellular Streptophyta, affiliated to Clausteriaceae and Desmidiaceae. The remaining archaeplastid OTUs affiliated to the chlorophyte classes Mamiellophyceae, Trebouxiophyceae, Nephroselmidophyceae and, mainly, Chlorophyceae. In addition, we detected a few OTUs belonging to other, less abundant eukaryotic groups. Thus, 4, 1 and 6 OTUs affiliated to katablepharids, telonemids and haptophytes, respectively (Fig. S3). Most haptophyte OTUs were very close to other freshwater sequences, but the OTU\_3295 had a sequence nearly identical to *Jomonlithus littoralis* (AM490979; 99% identity), a coastal species, and to two environmental sequences (JX680345, JX680344) from a brackish pond (Simon *et al.* 2013). Sequences belonging to the basal haptophyte group HAP-1 (Šlapeta *et al.* 2005) were detected in the Ru Sainte Anne (OTU\_3063; Fig. S3). Six additional OTUs belonged to centroheliozoans (Fig. S3). Their closest relatives were sequences from soil or freshwater sediment, and were only retrieved in Sainte Anne brook, the most narrow and less deep ecosystem (Table 1). That observation, along with the higher proportions and diversity in Sainte Anne brook of cercozoa, fungi and diatoms, known to be usually composed of large and/or benthic cells, could be explained by a higher influence or contribution of benthic communities in this system. One excavate OTU (OTU\_1857; 99% identical to *Trimastix marina*) and one apusomonad OTU were also detected in Sainte Anne brook. Three amoebozoan OTUs from the brook and 2 ancyromonad OTUs could also be detected in our small freshwater systems (Fig. S9). Finally, 66 of the total 812 OTUs determined could not be affiliated with confidence to any group.

*Effects of physicochemical parameters and distance on community structure and composition*

The five sampled ecosystems were separated by distances between 2 and 9.5 km (Fig. S1). To see whether the differences in community composition increased with geographical distance, we pooled sequence data from replicate samples for the 0.2-5 µm size fraction, calculated Bray-Curtis distances between all systems based on the pooled data, and performed a Mantel analysis to test the correlation between Bray-Curtis and geographical distance matrices. The Mantel test showed no correlation between differences in protist community composition and distance between ecosystems ( $r=0.2151$ ,  $P\text{-value}=0.236$ ). As an illustration, OTU\_52 (*Cryptomonas curvata*) reached 44% of all reads (43 - 52 % in replicates) in the Saint Robert pond while it could not be detected in the Gabard pond (Fig. 5), its nearest system (2.1 km away). However, this OTU was detected in all other systems, up to 9.5 km away.

As mentioned above, each of our five ecosystems was characterized by a set of specific environmental variables, with the Mare Gabard and the Ru Saint Anne being, respectively, the less and most charged in total dissolved solids and Saint Robert being highly enriched in ammonia (Table 1; Fig. 1). All these differences in environmental parameters correlated to the structure of small protists in these ecosystems as revealed by a Mantel test linking environmental and community pairwise distances between samples ( $r=0.7323$ ,  $P\text{-value}=0.014$ ). However, additional canonical correspondence analyses (CCA) did not reveal any clear relationship between any of the detected eukaryotic supergroups and one or several environmental variables, although some OTUs appeared to be associated with particular samples according to their specific presence (Fig. 5) and distribution in CCA (Fig. 7). At the phylum or class level (Fig. S10), haptophytes, katablepharids, choanoflagellates, cercozoans and cryptophytes appeared to reach higher proportions where pH was the highest and phosphate concentration the lowest, the two latter variables being tightly correlated in principal component analysis (PCA) and CCA (Figs. 1 and 7). Streptophytes were associated with high values of dissolved organic carbon while bacillariophytes and fungi were more abundant where total dissolved solutes were the highest. These features could thus partially explain the differences in composition of the small eukaryotes among our contrasting ecosystems.



**Figure 7: Canonical Correspondence Analysis (CCA) plot.** Only 0.2-5  $\mu\text{m}$  size fractions are considered. Each dot represents an OTU. The colors indicate the taxonomic affiliation. Black triangles indicate samples. Duplicate samples appear superimposed. CCTHK: Cryptophyta, Centroheliozoa, Telonemia, Haptophyta and Katablepharida. TDS: Total Dissolved Solutes, DOC: Dissolved Organic Carbon, OrthoP: orthophosphate.

## D- Discussion

### *Shallow freshwater systems, reservoirs of protist diversity*

Compared with oceans, molecular protist diversity surveys in freshwater ecosystems are still scarce. Paradoxically, inland water bodies are collectively much more heterogeneous than oceans, and much more sensitive to environmental variation. Despite so, only a handful of studies on freshwater protist communities exist, mainly from a variety of lakes, from temperate regions and mountains to polar areas (e.g. Charvet *et al.* 2012; Lepère *et al.* 2013; Taib *et al.* 2013). However, nearly nothing is known from smaller, shallower and often ephemeral freshwater systems, despite their global distribution and ecological importance (Downing *et al.* 2006). Recent approaches based on massive sequence analysis, though prone to a variety of methodological errors and biases, are called to facilitate comparative studies among those ecologically relevant ecosystems. This is especially true for tiny protists, since morphology-based exploration very often miss their phylogenetic diversity (Moreira & López-García 2002; Massana *et al.* 2002; Šlapeta *et al.* 2006). To contribute to that task, we carried out a study based on massive pyrosequencing of amplified 18S rRNA gene fragments of protists in the 0.2-5  $\mu\text{m}$  size range in five shallow freshwater ecosystems from a temperate area. Many of the high-rank taxa detected occur, although with variable relative abundances, in other freshwater systems, such as lakes. Cryptophytes, stramenopiles and alveolates were by far the most abundant supergroups detected, though their internal diversity varied greatly among ecosystems (Fig. 2). Although less abundant in the open sea (Shi *et al.* 2009; Kirkham *et al.* 2013), small cryptophytes seem to be common in freshwater (Lefranc *et al.* 2005;

Šlapeta *et al.* 2005; Lepère *et al.* 2008; Mangot *et al.* 2012; Taib *et al.* 2013). Stramenopiles, as well as ciliates, which dominate the alveolate diversity in our samples, are also common in protist communities from freshwater lakes (Lefranc *et al.* 2005; Lepère *et al.* 2008; Taib *et al.* 2013). Fungi, Ichthyosporea and choanoflagellates within the Opisthokonta, cercozoans within the Rhizaria, and members of the Amoebozoa, Archaeplastida, Excavata as well as members of the phylogenetically unresolved Haptophyta, Centrohelida, Katablepharida, Telonemida and Apusozoa were also detected.

Surprisingly, fungi, which are often detected in very large proportions in some lakes (Lefranc *et al.* 2005; Lepère *et al.* 2008), accounting for up to 94% of all the 454-pyrosequences in some of them (Taib *et al.* 2013), were in very low proportions (generally much less than 10%; Fig. 2) in our systems. This may reveal a real difference between shallow ecosystems and deeper lakes, but it could also reflect varying population dynamics along the year (Nolte *et al.* 2010). Our samples were collected in early spring and many of the fungal OTUs detected correspond to potential parasitic fungi or related lineages, such as chytrids, rozellids/cryptomycota or aphelids (Fig. S7), which might experience fluctuation depending on host population dynamics along the year. Temporal surveys would be needed to test such a hypothesis. Several OTUs in our freshwater systems branched deeply within known eukaryotic supergroups. Thus, the Rozellida-Cryptomycota and the aphelids have been considered the deepest lineages of fungi (Lara *et al.* 2010; Jones *et al.* 2011), although recently, they have been re-classified as forming part of the superphylum Opisthosporidia together with Microsporidia and would be the sister group to fungi (Karpov *et al.* 2014). Other detected deep-branching lineages include the Colpodellida (Leander *et al.* 2003) within the alveolates, the poorly known novel Clade-10 Cercozoa and Tremulida (Bass *et al.* 2009; Howe *et al.* 2011) or the early diverging haptophyte lineage HAP-1, so far only detected in freshwater systems (Šlapeta *et al.* 2005; Shalchian-Tabrizi *et al.* 2011).

Although protist community composition was different in the 5 analyzed systems, the comparison of protist diversity in two cell fraction sizes (0.2-5 µm and 5-30 µm) in the most diverse site, the Etang des Vallées, did not differ much in terms of high-rank taxa (Figs. 2 and 3). Indeed, despite the presence of OTUs apparently confined to each cell-size fraction (21 and 25%, for the smaller and larger size ranges, respectively), a considerable proportion of OTUs (53%) were shared between the two cell fractions (Fig. 5B). Several explanations are possible and non-mutually exclusive. One is that many small organisms may be retained in filters of larger pore-diameter if the filtered biomass is high, since they might be entangled with larger organisms. On the opposite, flexible organisms of relatively bigger sizes than those of filter pores may pass through them under the filtration pressure applied. Finally,

many organisms may encompass a size range that spans the 5  $\mu\text{m}$  diameter barrier imposed by our filters. This may be due to either differential sizes during their life cycle (e.g. gametes or spores and vegetative forms) or to size variability under a given life stage.

The broad protist diversity unveiled in the five freshwater systems also reflects a wide ecological diversity of functions. At a very general level, typical photosynthetic, heterotrophic and parasitic groups were detected. Among photosynthesizers, cryptophytes followed by photosynthetic stramenopiles (diatoms, chrysophytes, synurophytes, xanthophytes) largely exceeded green algae. Many of these lineages may also contribute to heterotrophic activities, since many photosynthetic protists are mixotrophs (Zubkov *et al.* 2008; Massana 2011; Hartmann *et al.* 2012). Typical predators span most of the eukaryotic diversity identified, from choanoflagellates and amoeba to bicosoecids, ciliates, cercozoa, centrohelids, katablepharids or apusozoa (Fig. 2). However, heterotrophic activities also encompass the degrading activity of osmotrophic taxa, including several fungi and labyrinthulids, and that of parasitic or parasitoid protists (many chytrids, cryptomycota, aphelids, oomycetes, apicomplexa, perkinsids).

Altogether, our extensive 18S rDNA-based survey of small protists suggests that shallow freshwater systems are important reservoirs of eukaryotic diversity. Not only members of all supergroups are present, but also several members of uncertain, poorly known or deep-branching lineages. Given the heterogeneity of this kind of systems on Earth, it can be advanced that further studies on shallow freshwater systems will uncover yet-to-characterize protists, the study of which will be of relevance to understand the ecology of these ecosystems and, from an evolutionary perspective, to reconstruct poorly resolved areas in the eukaryotic tree.

### *Marine-freshwater barriers transgressed*

One of the potential surprises that the study of protist diversity in varied freshwater systems may bring is the increasing awareness that salinity barriers can be overcome more easily than previously thought. Although lineages with members in both freshwater and marine systems exist, such as dinoflagellates, haptophytes, perkinsids or stramenopiles (Logares *et al.* 2007; Bråte *et al.* 2010b; Shalchian-Tabrizi *et al.* 2011; Simon *et al.* 2013), the marine-freshwater transition is thought to be rare (Logares *et al.* 2009). However, this view may be biased by current undersampling of freshwater systems. Indeed, in our study we have detected protist lineages thought to be exclusively marine in the past, notably several MAST lineages. Thus, we did not only identify members of MAST-2 and MAST-12 lineages, which have recently

been detected in other freshwater systems (Massana et al., 2013), but also members of MAST lineages never identified in freshwater systems before. This was the case of MAST-3, for which we identified two bona fide OTUs and, potentially, of MAST-6 and even MAST-1 lineages, although OTUs related to the latter are of much more uncertain affiliation (Fig. 6).

Moreover, in addition to the occurrence of the emblematic MAST groups in these shallow freshwater systems, we also identified many OTUs widespread in the eukaryotic tree that share 99% identity or more with sequences retrieved from marine environments. Examples are the haptophyte OTU\_3295, practically identical to *Jomonlithus littoralis* (Fig. S3), the stramenopile OTU\_116 (Fig. S4), the ciliate OTU\_255 (Fig. S6) or the excavate OTU\_1857 affiliating to *Trimastix marina* (Fig. S9). But there are many other OTUs whose closest relatives are sequences from marine systems, even if similarity is slightly lower.

In fact, even if salinity seems to be a relevant ecological determinant structuring microbial communities (Lozupone & Knight 2007), the marine-freshwater barrier seems to be relatively easy to cross. On the one hand, several protists, notably cryptophytes, are osmotolerant and can cope with various salinity concentrations by the means of contractile vacuole regulation (Hoef-Emden 2014). On the other hand, many freshwater systems contain relatively high levels of dissolved solutes (organic and/or inorganic) requiring similar adaptations as those require for life in seawater salts. In this sense, it is interesting to note that most of the "typically marine" lineages detected in our freshwater systems, were identified in the Ru Sainte Anne, which has the highest TDS content and seems greatly influenced by this parameter (Table 1; Fig. 7).

### *Elements of protist biogeography*

There are two essentially opposed views with regard to microbial and, more specifically, protist biogeography. Most classical views posit that small free-living protists would find little barriers to dispersal and, hence, be widely distributed. Differences in protist community structure would then be essentially explained by local environmental parameters. This "everything is everywhere, but the environment selects" view seems supported by the occurrence of cosmopolitan protist species (Baas-Becking 1934; Finlay 2002; Šlapeta *et al.* 2006). On the opposite extreme, a variety of studies seem to suggest that endemic protists (Foissner 2006) and taxa-area relationships exist for microbial eukaryotes (Green *et al.* 2004). Comparing community compositions sidesteps undersampling and the nearly impossible task of demonstrating true microbial endemisms (Martiny *et al.* 2006). A recent study suggested that the beta-diversity of the small eukaryotes between lakes was linked to the geographic distances between ecosystems (Lepère *et al.* 2013). However, it is extremely difficult to



disentangle the effect of local environmental parameters from physical distance. Furthermore, other factors, such as the temporal (Nolte *et al.* 2010) and the phylogenetic scale (Ragon *et al.* 2012) need to be taken into account. In our case, we sampled the five systems at the same or correlative dates to limit temporal effects and, although the distances involved were different from those in the study of Lepère *et al.* (2013) (10 km versus 100 km scale), our study clearly rejects geographic distance as a driver of community composition. On the contrary, Mantel tests show significant correlations between differences in community structure and physico-chemical parameters. However, clear associations between high taxon levels and environmental parameters were not identified (Fig. 7 and S10). Differences at finer, OTU scale might provide hints about the role of environmental selection in determining community structure. At any rate, testing which factors more greatly influence the composition of these communities would require the inclusion of biotic parameters, notably the diversity and relative abundance of bacteria, archaea and viruses from the same systems.

### E- Experimental procedures

#### *Sampling and measurement of physicochemical parameters*

Samples were collected in spring 2012 from five small and shallow freshwater ecosystems at the Natural Regional Park of the Chevreuse Valley (France, South of Paris) (Table 1 and Fig. S1). The systems were chosen to represent a variety of conditions at local scale, from forest ponds rich in organic matter (Mare Gabard) to more urban and agricultural-influenced systems (Saint Robert). Surface water was collected using sterile plastic carboys and processed immediately back in the laboratory. Water samples were serially filtered through 30 µm pore-size nylon filters (Millipore), and through 5 and 0.2 µm pore-size diameter Nucleopore membranes (Whatman). Filters were stored frozen at -20°C until DNA extraction. The water temperature, pH, the concentration of total dissolved solutes and the level of dissolved oxygen were measured *in situ* using a multiparameter probe (Multi 350i, WTW). The concentrations of dissolved nitrate ( $\text{NO}_3^-$ ), nitrite ( $\text{NO}_2^-$ ), ammonia ( $\text{NH}_3$ ), orthophosphate ( $\text{PO}_4^{3-}$ ) and dissolved organic carbon (DOC) were measured in water samples filtered through 0.2 µm pore-size diameter Nucleopore membranes on the same day of sampling using manufactured colorimetric tests (Hach-Lange). Chlorophyll *a* concentrations were determined after harvesting plankton biomass on glass microfiber filters (GF/F, Whatman) that were stored at -20°C until ethanol pigment extraction. For chlorophyll extraction, filters were dried then ground in 7 ml absolute ethanol, and heated at 70°C for 20 min. After centrifugation, 1 ml of supernatant was collected and optical densities at 665 and

750 nm were measured (spectrophotometer DR5000 Hach-Lange). Chlorophyll *a* concentration was determined on pigment extract by spectrophotometry, as follow:  $[Chla] = 11.9 * (OD_{665} - OD_{750}) / w * V_e / V_f$  with [Chla]: chlorophyll *a* concentration ( $\mu\text{g.l}^{-1}$ ), 11.95 :

Reciprocal specific absorbance coefficient of chlorophyll *a* at 665 nm ( $\mu\text{g.cm.ml}^{-1}$ ),  $OD_{665}$ : optical density at 665 nm,  $OD_{750}$ : optical density at 750 nm, *w*: width of the spectroscopic cuvette (1 cm),  $V_e$ : volume of the pigment extract (in ethanol, ml) and  $V_f$ : volume of water filtrated on the glass microfiber filter (l). The protocol was adapted from (Ritchie, 2006).

#### *DNA extraction, amplification and sequencing of 18S rRNA genes*

DNA was extracted from cells collected onto filters that were cut into pieces using the PowerSoil DNA extraction kit (MoBio) according to the manufacturer's instructions. DNA was eluted in 80  $\mu\text{l}$  of 10 mM Tris, pH 8.0. 18S rRNA gene fragments of ca. 550 bp, encompassing the V4 hypervariable region, were amplified using the newly designed primer EK-565F (5'-GCAGTTAAAAAGCTCGTAGT) and primer 18s-EUK- 1134-R - UNonMet (5'-TTTAAGTTTCAGCCTTGCG) biased against Metazoa (Bower *et al.* 2004). Both forward and reverse primers were tagged with 20 different 10 bp molecular identifiers (MIDs) to allow pooling and later differentiation of PCR products from 20 distinct samples. PCR amplifications were conducted in a total reaction volume of 25  $\mu\text{l}$  using 1.5 mM  $\text{MgCl}_2$ , 0.2 mM of each dNTP (PCR Nucleotide Mix, Promega), 0.3  $\mu\text{M}$  of each primer, 0.3 to 2  $\mu\text{l}$  of DNA sample and 0.5 U HotStart Taq polymerase (Taq Platinum, Invitrogen). The amplification conditions consisted of 25 cycles (94°C for 30 s, 58°C for 45 s and 72°C for 90 s), preceded by 3 min of denaturation at 94 °C and ending with a 10 min final extension step at 72°C. Amplicons from 5-7 independent PCR products for each sample were pooled together and then purified using the QIAquick PCR purification kit (Qiagen), according to the manufacturer's instructions. The same amounts (around 200 ng) of purified amplicons from 20 samples were pooled. Amplicons were pyrosequenced using the 454 GS FLX Titanium technology from Roche (Beckman Coulter Genomics).

#### *454 pyrosequence analysis*

We obtained a total of 265,899 pyrosequences (Table 1). A series of filters were applied in order to retain only high-quality sequences. First, pyrosequences containing errors in the primer region and positions with undetermined bases were eliminated using a local pipeline (Bachy *et al.* 2013). The remaining sequences were analyzed with AmpliconNoise (Quince *et al.* 2011) to further eliminate errors introduced during PCR reactions or 454 sequencing, and

build Operational Taxonomic Units (OTUs). Filtered reads were clustered by pairwise alignment and average linkage into OTUs with a 98% similarity cut-off using AmpliconNoise integrated in our local pipeline (Bachy *et al.* 2013). Singletons, i.e. OTUs composed of only one read, were eliminated for precaution. The most abundant sequence in each OTU was used as reference. OTU reference sequences were blasted against the Silva SSU111 database (Pruesse *et al.* 2007) and assigned to taxonomic groups based on sequence similarity. The sequences in all OTUs were then attributed to the different samples according to their MIDs. Chimerical OTUs were eliminated by a stringent procedure combining automated and manual steps. OTUs including sequences from at least two different samples were considered to be real. OTUs composed of sequences from only one sample were checked for chimeras using KeyDNATool (<http://KeyDNATools.com>). The sequences considered suspect by the software were double checked by comparing BLAST hits recovered from independent sequence fragments (sequences were split in 2 and 3 fragments). Finally, OTUs whose representative sequence had a coverage of less than 90% with its first BLAST hit were eliminated if they were present in only one sample; they were kept but with their taxonomic affiliation changed to “uncertain” if they were present in at least two samples. OTUs affiliated to cryptophyte nucleomorphs were excluded from our analysis. After filtering, we kept 146,549 correct reads (Table 1).

### *Phylogenetic analyses*

Phylogenetic trees were built for each eukaryotic super-group or for several high-rank taxonomic lineages if they comprised only a few OTUs. Analyses included representative sequences of OTUs, their first BLAST hit and sequences from the closest cultured members. Sequences were aligned using Probcons (Do *et al.* 2005). Positions retained to build trees were selected from the multiple alignments using Gblocks (Castresana 2000) with the less stringent parameters. Phylogenetic reconstructions were then carried out by maximum likelihood approximation using FastTree (Price *et al.* 2010). Trees were visualized using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). Sequences with particularly interesting positions in trees were then blasted against the NCBI database (<http://blast.ncbi.nlm.nih.gov>) to have an insight of their similarity with additional sequences in databases. The taxonomic indications given in trees are based on the taxonomic affiliation proposed in the PR2 database (<http://ssu-rrna.org/index.html>).

*Statistical analyses*

Statistical analyses were conducted using the R software (<http://cran.r-project.org>) (R Development Core Team 2011). To assess overall differences between microbial community compositions, we calculated pair-wise Bray-Curtis distances between all samples on the basis of OTU proportions (number of reads from each OTU for each sample normalized to the total number of reads in the corresponding sample) among replicates or sampling sites (Bray & Curtis 1957). This would avoid heterogeneity due to different numbers of sequences generated per sample. Non-metric multidimensional scaling (NMDS) ordination analyses were conducted based on Bray-Curtis distances (after applying a Wisconsin standardization to balance the influence of the most and least abundant OTUs) using the “Vegan” R package (Oksanen *et al.* 2013). Correspondence analysis (CoA) on all OTU frequencies for all 12 samples were done using the “Ade4” R package (Dray & Dufour 2007). Diversity and richness indices were determined using the “Vegan” package. Richness was estimated by rarefaction analysis as the estimated number of OTUs in a random subsample of each sequence library (raw counts of OTUs), of the same size as the smallest one (Hurlbert, 1971). Simpson index is defined as  $D' = 1 - \sum_{i=1}^S (f_i^2)$  (Simpson 1949) and evenness was calculated as  $e = -\sum_{i=1}^S f_i \ln(f_i) / \ln(S)$  (Pielou 1966) with S being the observed number of

OTUs and  $f_i$  the frequency of each OTU<sub>i</sub> in the sample. Venn diagrams showing the number of OTUs shared by, or exclusive to, the different samples, and heatmaps showing the presence/absence of OTUs were built using the “gplots” package (Bolker *et al.* 2012). To test whether community composition correlated with environmental parameters in the different samples, we constructed a matrix of Bray-Curtis dissimilarities based on OTU frequencies (sequences from replicate samples were pooled given that they clustered together in previous analyses) and a matrix of Euclidean distances based on physico-chemical parameters for all ecosystems using the “Vegan” package. Both matrices were compared using a Mantel test. The Bray-Curtis matrix was also compared to a matrix of geographical distances between ecosystems. Geographical distances were estimated based on coordinates ([http://biodiversityinformatics.amnh.org/open\\_source/gdmg/](http://biodiversityinformatics.amnh.org/open_source/gdmg/), using the mean earth radius (Moritz 2000) as spheroid. Principal Component Analysis (PCA) and Canonical Correspondence Analysis (CCA) were conducted on centered and scaled physico-chemical parameters and OTU frequencies in replicate samples (0.2-5 µm size fractions) using the “Ade4” package.

## Acknowledgements

We are thankful to F. Hardy, the Parc Naturel Régional de la Haute Vallée de Chevreuse and the Office National des Forêts du Parc de Rambouillet. We also thank Giselle Walker for critical reading of the manuscript. The research leading to these results received funding from the CNRS EC2CO program and the European Research Council under the European Union's Seventh Framework Program ERC Grant Agreement 322669 'ProtistWorld'.

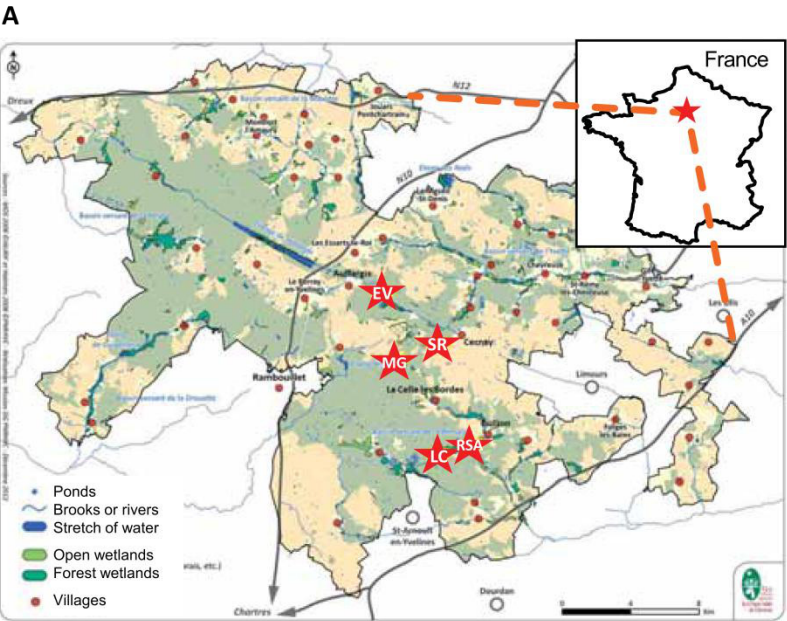
## F- Supplementary Information

**Note:** Supplementary information also contained phylogenetic trees for all supergroups, but they were too large to be readable once printed. They are only included in the pdf version of the manuscript, as an appendix.

Table S1: Number and percentage of reads assigned to different taxa in the five studied ecosystems . Replica samples are labelled 'b'.

Sampled ecosystem	Gabard pond		Saint Robert pond		Etang des Vallées			Sainte Anne brook		La Claye pond	
	0.2 - 5 µm		0.2 - 5 µm		5 - 30 µm		0.2 - 5 µm		0.2 - 5 µm		
Size fraction	MG25	MG25b	SR25	SR25b	EV34	EV34b	EV33	EV33c	RSA25	RSA25b	LC25
Sample names	10616		42034	4506	10982	6906	17670	3947	19652	4243	11191
<div><div></div><div>Nbr of reads per sample</div><div></div><div>Nbr of reads per group</div></div>											
Fungi	2958	2.6	3.5	0.1	0.2	0.9	0.3	0.3	8.6	11.9	0.7
Opisthokonta	30	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.1
Ichthyosporia	16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Choanoflagellida	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Discosea	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amoebozoa	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Excavata	3081	4.7	7.0	0.0	3.1	5.6	3.0	2.1	3.2	2.9	0.6
Chlorophyta	115	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.2	0.3	0.2
Streptophyta	368	0.0	0.0	0.0	0.0	0.6	0.6	0.8	0.5	0.6	0.0
MAST	757	2.3	1.2	0.0	0.1	0.5	0.2	1.6	0.0	0.0	0.4
Bicosoecida	8473	0.0	0.0	0.0	0.0	13.1	3.5	4.4	27.1	13.3	0.0
Bacillariophyceae	1471	0.0	0.1	0.0	0.0	0.1	0.0	0.1	6.1	4.8	0.0
Oomyceta	11872	6.2	3.4	8.2	2.2	14.2	12.7	14.9	5.9	5.1	2.9
Chrysophyceae	7964	20.1	32.0	4.3	2.1	7.3	5.7	3.7	1.4	0.9	1.2
Synurophyceae	14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
Labyrinthulida	56	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.0
Xanthophyceae	5541	1.8	1.0	0.0	0.4	3.0	1.9	5.8	0.9	0.8	19.1
Other Stram	23037	36.2	16.1	12.1	19.8	21.8	33.9	28.8	2.9	1.3	3.1
Ciliophora	730	3.7	4.6	0.0	0.0	0.6	0.3	0.3	0.0	0.0	0.0
Dinoflagellata	14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
Alveolata	362	0.5	0.8	0.0	0.0	0.3	0.2	0.3	0.3	0.3	0.3
Perkinsea	1273	0.0	0.0	0.0	0.7	4.9	2.0	2.2	0.2	0.1	0.1
Other Alveolata	1829	0.1	0.1	0.0	0.1	1.0	1.0	0.6	5.9	7.5	0.3
Cercozoa	24	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0
Haptophyta	71564	21.9	29.9	75.2	71.4	19.1	25.3	22.8	31.7	43.7	69.9
Cryptophyta	38	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0
Centroheliozoa	3630	0.0	0.0	0.0	0.1	6.7	9.2	10.9	0.1	0.0	0.0
Katablepharida	18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Telonemida	17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
Ancyromonadida	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Apusomonadidae	1284	0.0	0.2	0.0	0.0	0.1	0.1	0.2	4.2	5.6	1.1
Uncertain											

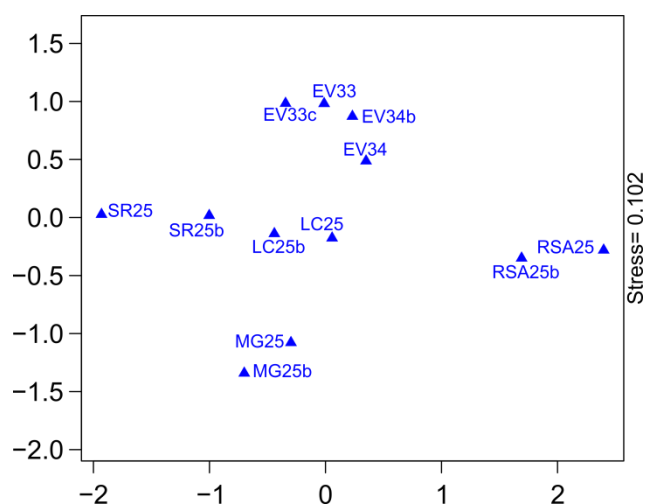




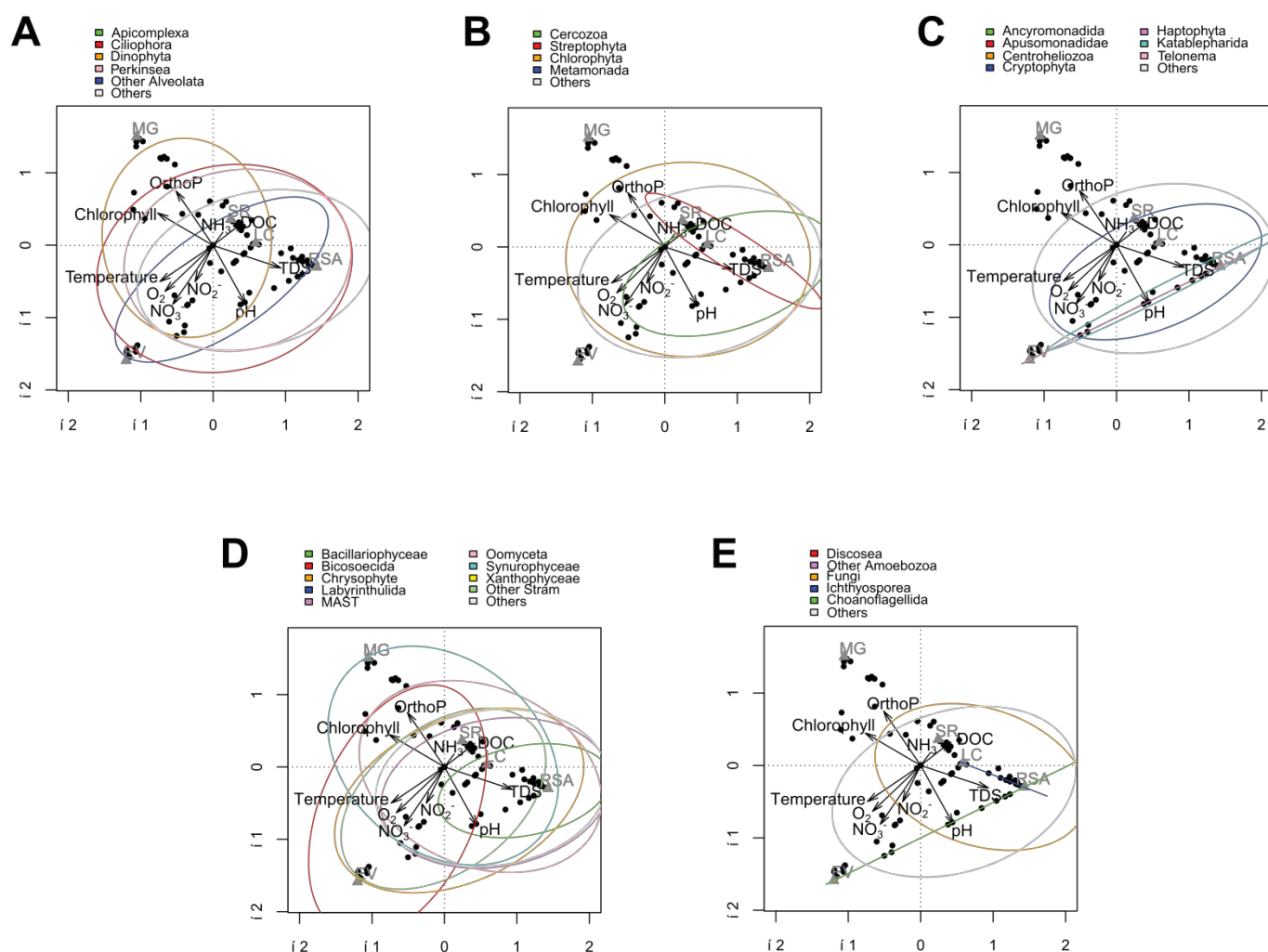
B

	Gabard	Saint Robert	Etang des Vallées	Sainte Anne	La Claye
Gabard	0	2.11	3.95	5.87	5.2
Saint Robert		0	3.48	6.13	6.30
Etang des Vallées			0	9.46	9.14
Sainte Anne				0	2.48
La Claye					0

**Figure S1.** Shallow freshwater systems sampled in this study in April 2011. A. Localization of the five ecosystems in the Parc Natuel Regional de la Haute Vallée de Chevreuse, SouthWest of Paris, France. B. Distance as the crows flies between the sampled ponds and brook (km). Photographs of the sites are shown on the left.



**Fig. S2.** Nonmetric Multi-Dimensional Scaling (NMDS) plot showing protist community composition similarities and differences among the five ecosystems studied. MG25 and MG25b: Mare Gabard, SR25 and SR25b: Saint Robert, EV33 and EV33c: Etang des Vallées (0.2-5  $\mu\text{m}$  size sample), EV34 and EV34b: Etang des Vallées (5-30  $\mu\text{m}$  size samples), RSA25 and RSA25b: Sainte Anne, LC25 and LC25b: La Claye



**Fig. S11.** Canonical Correspondence Analysis (CCA) plot. All plots are from the same analysis as in Fig.7, with ellipses emphasizing the distribution of alveolates (A), rhizarians, archaeplastidans and excavates (B), cryptophytes, centroheliozoans, telonemids, haptophytes, katablepharids and apusozoans (C), stramenopiles (D) and unikonts (E). Projected inertia on CCA1: 34,07% and on CCA2: 29.77%. Black dots correspond to OTUs. Grey triangles indicate the samples. SR, Saint Robert; EV, Etang des Vallées, MG: Mare Gabard; LC, La Claye; RSA, Ru Sainte Anne. Replicate samples are superimposed. DOC, Dissolved Organic Carbon; OrthoP, Ortho Phosphate; TDS, Total Dissolved Solutes.

### III. Epilogue

#### A- Une surprenante diversité des MAST en eau douce

Les MAST (MARine STRamenopile) forment un ensemble polyphylétique de taxa qui doit son nom à sa découverte en milieu marin, par des méthodes moléculaires (clonage et séquençage Sanger d'amplicons d'ADNr 18S) (Diez *et al.* 2001; Moon-van der Staay *et al.* 2001; López-García *et al.* 2001; Massana *et al.* 2004b). Par la suite, il a essentiellement continué à être détecté en milieu marin, dans divers océans et mers du monde entier (Massana *et al.* 2013). Cependant, des séquences affiliées aux lignées MAST-2 et MAST-12 ont également été détectées en eau douce. Jusqu'à ce jour, aucune séquence affiliée à un autre groupe marin n'a pu être identifiée en eau douce, et aucune lignée n'est exclusivement rencontrée en eau douce (Massana *et al.* 2013).

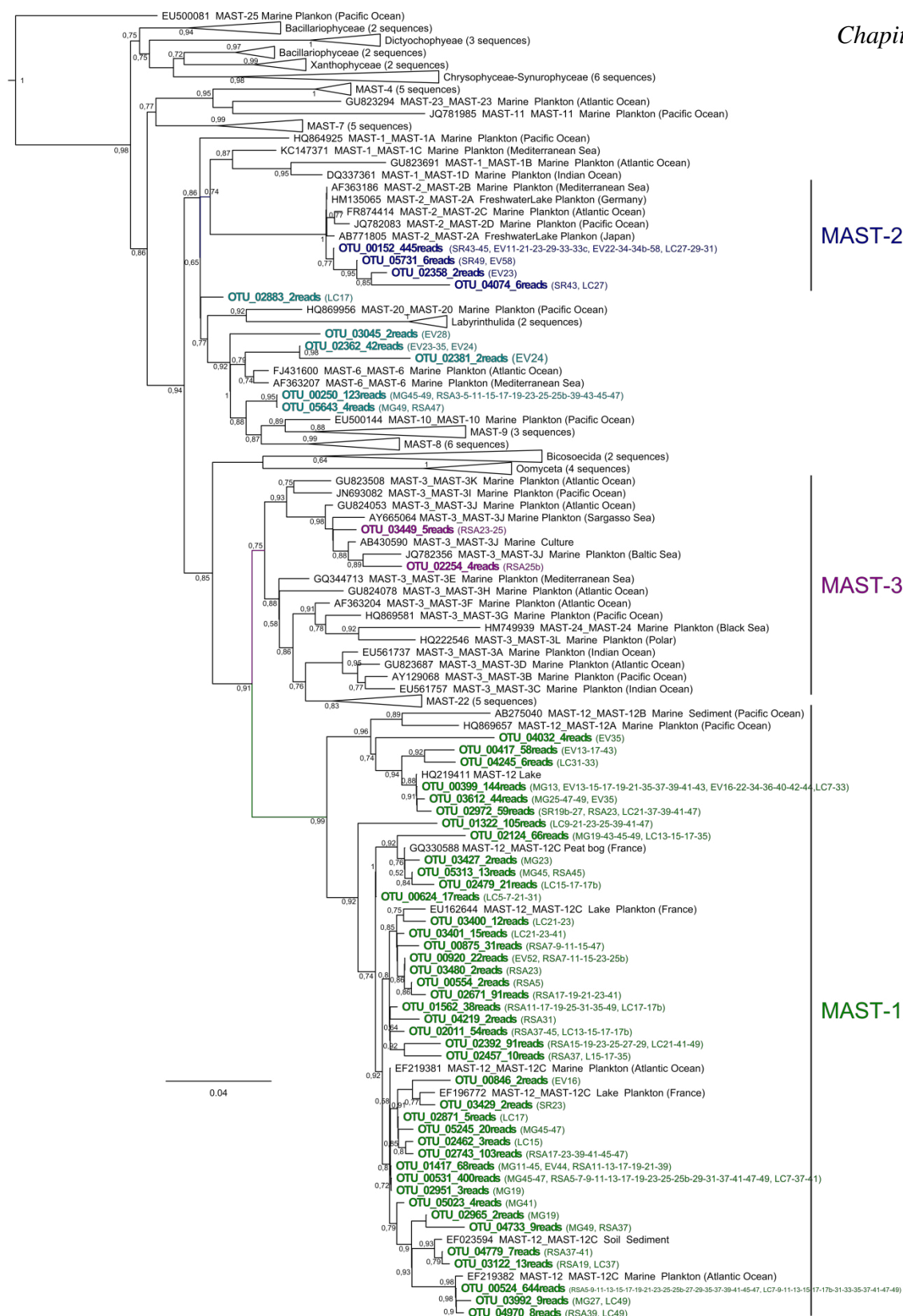
Nous avons détecté dans les échantillons prélevés au cours du mois d'avril 2012 douze OTU affiliées aux groupes MAST. Neuf et une OTU étaient affiliées aux groupes dont la présence en eau douce était déjà connue, MAST-12 et MAST-2, respectivement (Cf article). Cependant, nous avons également pu détecter deux OTU affiliées clairement à la lignée MAST-3, jusqu'alors exclusivement marin, et deux OTU relativement proches des groupes MAST-6 et MAST-1, mais dont les relations avec les différents groupes de MAST et les autres straménopiles étaient plus incertaines.

Les cinq mêmes écosystèmes que ceux explorés en avril 2012 ont été échantillonnés mensuellement pendant deux ans (Cf. Matériel et Méthodes général). Au cours de ce suivi, de nouvelles OTU affiliées à des groupes MAST ont pu être détectées. Les séquences représentatives de toutes les OTU MAST détectées toutes dates, tous écosystèmes et toutes fractions de taille confondues ont été alignées avec les séquences utilisées pour la figure 5 de l'article. L'arbre phylogénétique issu de cet alignement est présenté en figure 8.

On observe tout d'abord un plus grand nombre d'OTU affiliées aux clades MAST-2 et MAST-12 qu'en avril 2012. La plupart des OTU MAST-12 branchent dans le sub-groupe MAST-12C, avec les séquences issues d'eau douce et de sol (Massana *et al.* 2013). Cependant, 6 autres OTU apparaissent plus proches des groupes MAST-12A et 12B.

Nous n'avons pas détecté au cours des 24 mois de suivi plus d'OTU affiliées à MAST-3 qu'en avril 2012, mais cette nouvelle reconstruction phylogénétique confirme leur inclusion claire au sein de ce groupe présumé exclusivement marin.





**Figure 8.** Arbre phylogénétique en maximum de vraisemblance approximée des straménopiles, montrant la diversité des groupes MAST. 367 positions informatives ont été utilisées pour reconstruire la phylogénie. 2 séquences de ciliés ont été utilisés pour raciner l'arbre. Les séquences issues de ce travail sont indiquées en gras et en couleur. Les bootstraps supérieurs à 50% sont indiqués aux nœuds. Les échantillons à partir desquels les fragments d'ADNr 18S de MAST ont pu être amplifiés sont indiqués après le nom de chaque OTU, ainsi que le nombre de reads formant chaque OTU.

Enfin, 6 OTU supplémentaires sont affiliées par similarité de séquences (BLAST) aux MAST, mais leur position dans l'arbre des straménopiles est incertaine. (Figure 8, OTU en turquoise). Les séquences représentatives de ces OTU partagent seulement 91% à 96,5% d'identité avec leur première correspondance par BLAST, et paraissent donc assez distantes des séquences déjà référencées. Deux d'entre elles (OTU 2362 et 2381) forment un groupe monophylétique relativement soutenu avec les deux séquences de MAST-6 incluses dans l'analyse. Ainsi, en plus de MAST-3, d'autres groupes pourraient contenir des séquences issues de milieu d'eau douce. Il n'est pas impossible que certaines de ces OTU divergentes constituent un ou des nouveaux groupes, qui seraient alors exclusivement connus en eau douce. Cependant, avant toute spéculation, des séquences plus longues que celles obtenues par pyroséquençage 454 sont indispensables pour préciser la position de ces OTU dans l'arbre des straménopiles et plus précisément leurs relations avec les groupes MAST.

Quoi qu'il en soit, l'exploration de petits écosystèmes d'eau douce de petite taille nous aura permis de révéler la relativement grande diversité en eau douce d'un ensemble de taxons très abondants et répandus en milieu marin. Comme pour les haptophytes, ils paraissent moins abondants en eau douce que dans les mers et océans (0.27% des reads par échantillons en moyenne, moyenne calculée sur l'ensemble des échantillons et réplicats). Ils sont également moins divers mais de futures explorations de nouveaux écosystèmes pourraient révéler la présence d'une grande diversité encore inconnue.

#### **B- Des communautés très proches dans des échantillons correspondant théoriquement à des fractions de taille cellulaire différentes.**

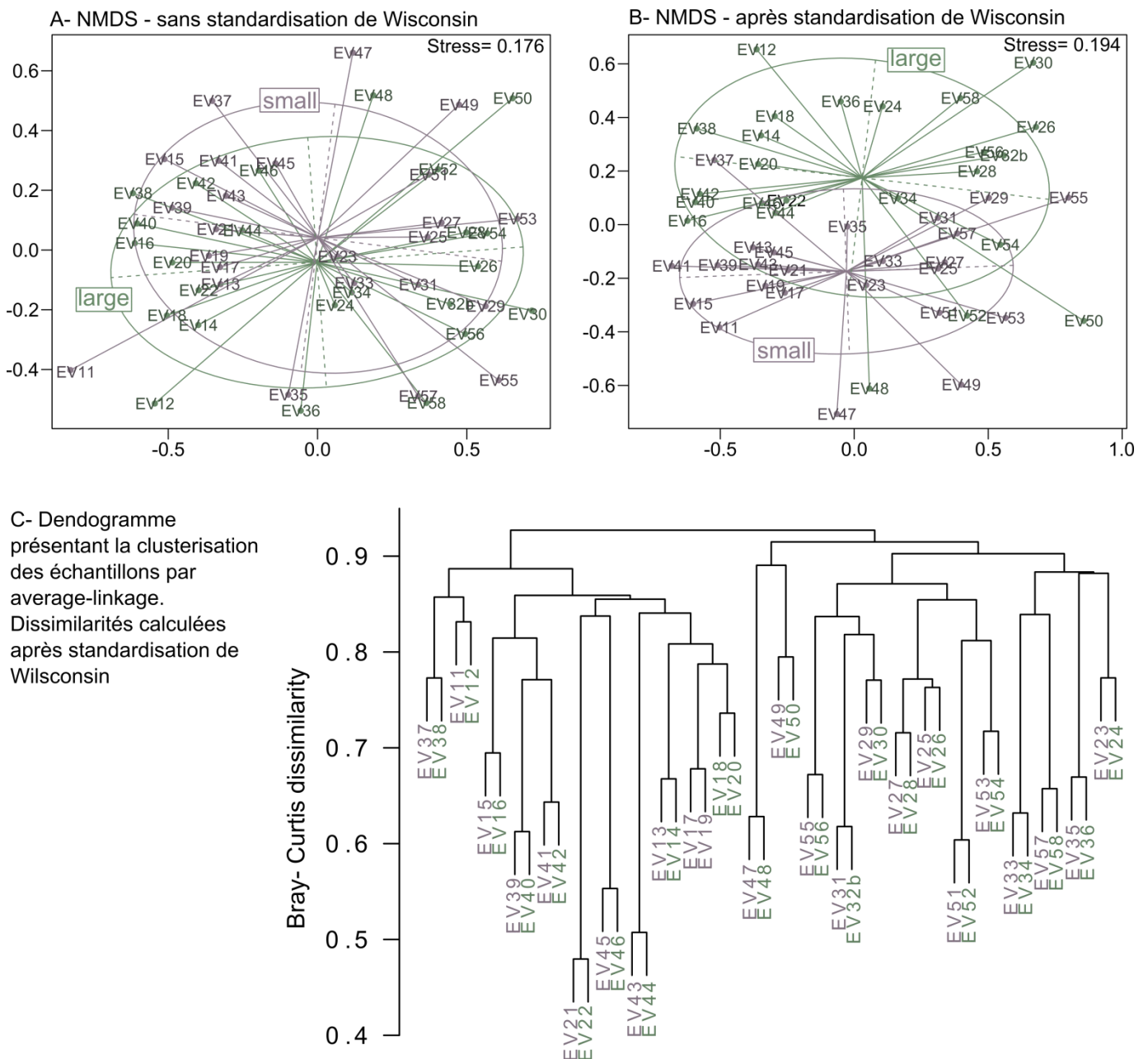
Pour notre analyse de la diversité des petits eucaryotes dans quatre mares et un ruisseau, les échantillons d'eau de surface ont été préfiltrés sur des filtres de 5  $\mu\text{m}$  de porosité avant que les cellules planctoniques ne soient recueillies sur des filtres de porosité 0,2  $\mu\text{m}$ . De cette façon, nous avons donc théoriquement sélectionné les cellules dont le diamètre est compris entre 0,2 et 5  $\mu\text{m}$ . Pour l'étang des Vallées uniquement, nous avons également recueilli sur les filtres de 5  $\mu\text{m}$  les cellules de diamètre compris entre 5 et 30  $\mu\text{m}$ , grâce à une préfiltration des échantillons sur des filtres de 30  $\mu\text{m}$ . Nous avons choisi de cibler également cette fraction de taille dans l'étang des Vallées, parce que nous y avons précédemment détecté une grande partie des OTU d'haptophytes formant le nouveau groupe EV d'isochrysidales (Cf. chapitre 1).

Dans le travail qui fait l'objet du précédent article, les librairies de séquences correspondant aux deux fractions de taille prélevées dans l'étang des Vallées au mois d'avril 2012 ont été comparées. Les communautés détectées dans ces deux échantillons se sont révélées très semblables, notamment au niveau de la répartition des grands groupes taxonomiques (Cf. article).

Afin de déterminer si cette similitude était propre aux échantillons d'avril 2012 (un problème ponctuel de manipulation lors des filtrations pourrait être en cause), nous avons également comparé les assemblages d'eucaryotes microbiens des fractions de taille 0,2-5  $\mu\text{m}$  et 5-30  $\mu\text{m}$  échantillonnés mensuellement lors des deux ans de suivi dans l'étang des Vallées.

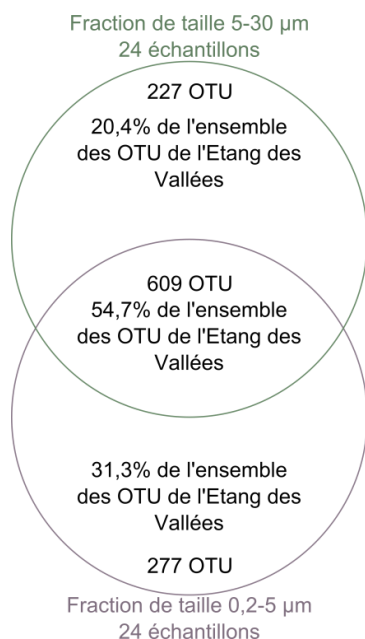
Dans un premier temps, les indices de dissimilarité de Bray-Curtis ont été calculés au sein de toutes les paires d'échantillons de l'étang des Vallées, des deux fractions de taille. Ces dissimilarités sont représentées dans un plan par NMDS (Figure 9A). Les ensembles d'échantillons des deux fractions de taille sont presque parfaitement superposés, indiquant que les communautés détectées dans l'une des fractions de tailles ne sont pas plus proches entre elles que des assemblages détectés dans l'autre fraction. Cependant, les indices de dissimilarité de Bray-Curtis calculés directement sur les fréquences des OTU dans les échantillons sont fortement influencés par les OTU les plus abondantes. J'ai donc également calculé les indices de dissimilarité de Bray-Curtis après double standardisation de Wisconsin des données. Le graphique NMDS correspondant (Figure 9B) montre une plus grande séparation des groupes d'échantillons selon la fraction de taille, bien que plusieurs échantillons de chaque fraction soient plus proches de l'autre. Ainsi, il semblerait que les échantillons des deux fractions de tailles contiennent des communautés très semblables du point de vue des OTU abondantes, mais diffèrent un peu plus par les OTU moins abondantes. Cependant, une clusterisation des échantillons par average-linkage, basée sur les indices de Bray-Curtis calculées après standardisation des données indique que l'échantillon le plus proche de chaque échantillon de la fraction 0,2-5  $\mu\text{m}$  est toujours l'échantillon de la fraction 5-30  $\mu\text{m}$  du même mois (Figure 9C). Seuls les échantillons de fin juillet et début septembre se regroupent d'abord selon la fraction de taille.





**Figure 9.** Comparaison des échantillons de cellules des fractions de taille 0,2-5  $\mu\text{m}$  (numéros impairs, en violet) et 5-30  $\mu\text{m}$  (numéros pairs, en vert) issus de l'étang des Vallées. Les échantillons sont comparés sur la base des dissimilarités de Bray-Curtis sans (A) ou après standardisation préalable des données (B et C). A et B : Non-metric MultiDimensional Scaling, C : clusterisation.

De plus, près de 55% des OTU détectées dans l'étang des Vallées au cours des 24 mois de suivi ont été détectées à la fois dans la fraction de taille 0,2-5  $\mu\text{m}$  et 5-30  $\mu\text{m}$  (Figure 10). Elles correspondent donc à des organismes qui peuvent à la fois être retenus et traverser les filtres de 5  $\mu\text{m}$  de porosité. La même proportion d'OTU partagées avait été estimée sur les échantillons prélevés en avril 2012 (53% Cf. article).



*Figure 10. Diagramme de Venn présentant le nombre d'OTU partagées ou spécifiques des fractions de taille 0,2-5 µm et 5-30 µm. Les séquences détectées dans les 24 échantillons de chaque fraction de taille ont été poolées.*

Il semblerait donc que la similarité entre les deux fractions de taille remarquée avec les échantillons prélevés en avril 2012 soit un fait général, observable sur un grand nombre d'échantillons. De nombreuses causes pourraient expliquer cette observation. Tout d'abord, la méthode de tri par filtrations successives n'est très certainement pas parfaite. Des cellules de petite taille pourraient rester sur le filtre de 5 µm de porosité, retenues par des cellules de plus grande tailles ou des particules qui colmatent en partie les pores du filtre. Au contraire, des cellules de grande taille pourraient se déformer sous l'effet de la pression négative appliquée sur les filtres et passer ainsi des pores plus petits qu'elles. Mais ces ressemblances entre fractions de tailles différentes pourraient également être dues à la variabilité de taille au sein d'une même espèce. En effet, une espèce dont la taille moyenne est aux alentours de 5 µm pourrait grouper des organismes de taille un peu inférieure et un peu supérieure à 5 µm, qui passeraient ou non le filtre, mais dont les gènes d'ARNr 18S se regrouperaient en une même OTU. Il est également possible qu'un même organisme passe par des stades de différente taille au cours de sa vie (e.g. stades haploïde et diploïde de taille différente). Enfin, une partie des OTU communes aux deux fractions de taille pourraient correspondre à des cellules de diamètre supérieur à 5 µm mais dont l'ADN pourrait passer les filtres après lyse des cellules. Une observation microscopique des cellules recueillies sur chaque filtre pourrait peut-être permettre de commencer à appréhender la part relative de ces différentes explications.



# Chapitre 3 : Diversité, écologie et dynamique des communautés de petits eucaryotes au cours d'un suivi mensuel sur deux ans dans quatre mares et un ru.

---



*"It has long been an axiom of mine that little things are infinitely the most important"*  
Sherlock Holmes (Propos retranscrits par Sir Arthur Conan Doyle, *A Case of Identity*).



## I. Introduction : Contexte et objectifs

Dans le chapitre précédent, nous avons pu mettre en évidence l'existence d'une grande diversité de petits eucaryotes dans quatre mares et un ru représentant la variété des petits écosystèmes d'eau douce. Nous avons également observé que la composition des communautés de petits eucaryotes était nettement influencée par les caractéristiques physico-chimiques de leur écosystème.

Or, ces caractéristiques sont susceptibles d'être extrêmement variables dans les écosystèmes aquatiques de petite taille tels que les mares et ruisseaux. En effet, le petit volume d'eau qu'ils contiennent n'a qu'un pouvoir tampon limité. Ainsi, les variations de température de l'air sont rapidement répercutées dans toute la colonne d'eau. L'arrivée, même en quantité limitée, de nutriments ou de polluants depuis les berges, ou la dilution due à de fortes pluies, peuvent également entraîner des variations non négligeables de la concentration de ces composés. Ces écosystèmes sont facilement brassés sur tout leur volume par le vent, ce qui entraîne des interactions régulières entre sédiments tapissant le fond de l'écosystème et l'eau jusqu'en surface. Ainsi, les petits écosystèmes d'eau douce peu profonds sont susceptibles d'être très dynamiques suite à des variations saisonnières ou ponctuelles de leur environnement proche.

Plusieurs études ont montré que les assemblages d'eucaryotes microbiens étaient très dynamiques (e.g. Nolte *et al.* 2010; Mangot *et al.* 2012; Kim *et al.* 2014), et notamment dans des mares (Bamforth 1958; Lara *et al.* 2011). Nous avons alors voulu décrire la dynamique des communautés de petits eucaryotes planctoniques, en parallèle des variations des caractéristiques physico-chimiques de l'eau.

Pour le travail présenté dans ce chapitre, nous avons pour principaux objectifs de :

- Poursuivre la description de la diversité des petits eucaryotes planctoniques dans les écosystèmes peu étudiés que sont les mares et ruisseaux.
- Décrire la dynamique des assemblages de petits eucaryotes dans ces écosystèmes très variables. Notamment, nous voulions savoir si cette dynamique présentait une saisonnalité, puisque l'existence d'une saisonnalité des communautés microbiennes et en particulier eucaryotes n'est pas claire.
- Identifier d'éventuels paramètres physico-chimiques potentiellement impliqués dans la dynamique de certains taxa.



**II.       Manuscrit de l'article 3 : Hectic temporal dynamics of complex  
small eukaryote communities in shallow freshwater systems**

**Hectic temporal dynamics of complex small eukaryote communities in  
shallow freshwater systems**

Marianne Simon, Purificación López-García, Philippe Deschamps, David Moreira, Gwendal Restoux, Paola Bertolino, Ludwig Jardillier

*Manuscript in preparation*

## A- Abstract

Small eukaryotes play key roles in the functioning of aquatic ecosystems, through their impact on both their biotic and abiotic close environment and, at a larger scale, on biosphere, by influencing climate. A few discrete studies revealed their high molecular diversity in small and shallow freshwater ecosystems. However, temporal and spatial dynamics of small eukaryote assemblages remain underexplored in these systems despite their tremendous abundance, wide distribution and rapid dynamics. Here, we conducted a monthly two-year survey of the small eukaryote community composition based on massive 454-pyrosequencing of amplified 18S rDNA fragments, in four ponds and one brook. Temporal survey allowed the detection of a broad diversity of eukaryotes, with 3 742 stringently defined phylotypes encompassing all recognized supergroups and taxa of uncertain affiliation, including low-abundant groups noticeable at a few dates only. A restricted proportion of the phylotypes were shared by the five systems, most of them being specific to one environment. Communities detected from the five geographically close ecosystems were highly different. Their dynamics was rapid, and strong signals of seasonality were observed. Low frequent OTUs composed the vast majority of the small eukaryote community. Among those detected several times, some of them remained at low abundance throughout the whole survey while some others occasionally reached relatively high abundances. Some of those low frequent taxa might constitute a seed bank and contribute to the stability of the ecosystem functioning. We identified potential ecological traits for fungi, ciliates, chlorophytes and cryptophytes but a large part of the ecology of this microbial community remained unexplained.

### Keywords:

18S rDNA / dynamics / freshwater / microbial eukaryotes / plankton / ponds

## B- Introduction

Protists, and small eukaryotes in particular (defined as cells  $< 5 \mu\text{m}$  in size herein), are ecologically important because of their multiple roles in the trophic food webs. Indeed, they belong to both heterotrophic (e.g. predators, decomposers and parasites) and photosynthetic functional groups, many being also mixotrophic (Zubkov & Tarran 2008). In addition, they are involved in several biogeochemical cycles at local and large scales (Li 1994; Cloern 1996; Jordan & Chamberlain 1997; Jardillier *et al.* 2010) or in climate regulation (Simó 2001), and toxic species can trigger massive death of other community members (Scholin *et al.* 2000; Edvardsen & Imai 2006). The high molecular diversity of small eukaryotes was first revealed in the oceans more than a decade ago (Diez *et al.* 2001; Moon-van der Staay *et al.* 2001; López-García *et al.* 2001). Since then, numerous studies have been conducted in marine environments, continuously discovering new groups of protists (Massana *et al.* 2004b; Shi *et al.* 2009; Kim & Harrison 2011).

That great microbial diversity may contribute to the resilience and explain the stability of the functioning of ecosystems despite shifts in environmental conditions (Shade *et al.* 2012a). Indeed, small eukaryote assemblages react to temporal variability of environmental conditions and biotic factors by a permanent reassembling over time, in marine systems and lakes (Lepère *et al.* 2006; Lara *et al.* 2011; Mangot *et al.* 2012; Jones *et al.* 2013). Those changes in community composition have been observed at all time-scales (Bamforth 1958; Cloern 1996; Vigil *et al.* 2009; Nolte *et al.* 2010; Mangot *et al.* 2012). The existence of a seasonality repeatable over years is not clear (Lepère *et al.* 2006; Jones *et al.* 2013; Kim *et al.* 2014), especially due to the scarcity of survey longer than one year. There is thus a need for pluriannual studies of small-eukaryote community dynamics. Over all time-scales, eukaryote assemblage dynamics appears driven by regional (e.g. pluviometry, light intensity) and local environmental parameters changes. Both top-down (predation, viral lysis) and bottom-up effects (e.g. prey or inorganic nutriment availability) are suspected to control the dynamics of microbial assemblages in all type of aquatic systems (Kalff & Knoechel 1978; Carrias *et al.* 1998; Hitchman & Jones 2000; Lepère *et al.* 2006; Lara *et al.* 2011; Zhao *et al.* 2011). Physical characteristics of water such as pH and temperature are also associated with community composition variability, but, altogether, parameters taken into account in environmental studies often explain only a small part of community dynamics (Lepère *et al.* 2006; Kim *et al.* 2014). Altogether, our knowledge of the small-eukaryote dynamics across space and time remains restricted.

Eukaryotic diversity started to be investigated by molecular surveys in freshwaters (Lefranc *et al.* 2005; Richards *et al.* 2005; Šlapeta *et al.* 2005) later than in oceans, and those

surveys remain, by far, less numerous (Zinger *et al.* 2011). If large lakes have been the most studied, small and shallow freshwater ecosystems (e.g. ponds and brooks) have barely been considered though they are highly numerous and thus represent a large part of continental surface covered by freshwaters (Downing *et al.* 2006; Downing 2010). Lakes and small ecosystems differ significantly in their functioning, because of the higher importance of the benthic and littoral zones in the latter (Søndergaard *et al.* 2005). Small systems are also characterised by a larger gradient of biotic and environmental parameters than lakes and rivers (Céréghino *et al.* 2008). In addition, species-richness of phytoplankton and of several macro-organisms per unit area seems to increase when the aquatic ecosystem area decreases (Downing 2010). A few molecular studies conducted on such small ecosystems revealed a wide diversity of small eukaryotes (Šlapeta *et al.* 2005; Lara *et al.* 2011; Simon *et al.* accepted). Because of these first observations and all these unique features, small and shallow freshwater ecosystems are expected to host widely diverse small-eukaryote communities, whose composition and ecological functioning might differ from the one in large lakes. In particular, ponds and brooks are highly dynamic (Bamforth 1958; Angélibert *et al.* 2004), as they are very sensitive to environmental modifications due to their low buffering capacity. Meteorological fluctuations, of ambient temperature or sunshine intensity for example, can rapidly be reflected in the entire water column, and can lead to the drying of the system. Moreover, nutrient concentrations can quickly rise by leaching of ridges during strong rain episodes, and small and shallow aquatic systems could also be more sensitive to anthropic effects (e.g. fertilizer or weed killer release). So, microbial assemblages hosted by such systems must respond to both strong environmental seasonal variations and occasional events. Actually, a few studies conducted in small ponds or a peat bog pool highlighted important variations of eukaryotic community compositions, linked to several measured environmental characteristics (Bamforth 1958; Hitchman & Jones 2000; Lara *et al.* 2011). Although understudied, small and shallow freshwater ecosystems thus appear to constitute adequate environments to understand the dynamics of the small-eukaryote communities. They are also particularly suited to describe the spatial distribution of microbial eukaryote assemblages, due to the presence of highly diverse systems in restricted geographical areas. We thus conducted a two-year monthly survey of the small-eukaryote community composition based on 454-pyrosequencing of the 18S rRNA gene in five small and shallow ecosystems including four ponds and one brook. Several environmental parameters were also measured and their potential influences on communities were deciphered thanks to multivariate statistical analyses. This is the first study to our knowledge that contributes in this way to the investigation of the molecular diversity and ecology of small eukaryotes in such systems.

## C- Material and Methods

### *Sampling and measurement of physicochemical parameters*

Samples were collected monthly from April 2011 to April 2013, in four small and shallow ponds and one brook, all located in a small geographic area, in the Natural Regional Park of the Chevreuse Valley (South of Paris, France) and characterised by a large variety of local environmental conditions (Supplementary Table 1, Simon *et al.*, accepted). Surface water was sampled using sterilized plastic carboys and processed a couple of hours later, immediately back in the laboratory. Samples were prefiltered through 5 µm pore-size Nucleopore membranes (Whatman) and small planktonic cells were collected onto 0.2 µm pore-size Nucleopore membranes. Membranes were immediately stored frozen at -20°C until DNA extraction. Water temperature, conductivity, pH and the concentration of dissolved oxygen were measured *in situ* using a multiparameter probe (Multi350i, WTW). The concentrations of dissolved orthophosphate ( $\text{PO}_4^{3-}$ ), organic carbon (DOC), nitrate ( $\text{NO}_3^-$ ), nitrite ( $\text{NO}_2^-$ ) and ammonia ( $\text{NH}_3$ ) were measured the same day as sampling, in water filtered through 0.2 µm pore-size Nucleopore membranes, using manufactured colorimetric tests (Hach-Lange). Chlorophyll *a* concentration was determined by spectrometry after ethanol pigment extraction from plankton biomass harvested on GF/F filters (Whatman), as previously described in Simon *et al.*, (accepted). Pluviometry, mean daily radiation and mean air temperature of the week before sampling, as well as the mean wind speed measured at 10 m high during the sampling day, were obtained from MétéoFrance (<https://donneespubliques.meteofrance.fr/>). Pluviometry was obtained by averaging the data recorded at the three closest meteorological stations: Choisel (48°41'00"N-2°00'00"E), le Perray (48°41'36"N-1°52'00"E) and Saint Arnoult (48°34'18"N-1°55'54"E) while radiation, temperature and wind speed were measured at station Trappes (48°46'24"N-2°00'30"E).

### *DNA extraction, amplification and sequencing of 18S rDNA fragments*

DNA was extracted using the PowerSoil DNA extraction kit (MoBio) from biomass collected onto 0.2 µm pore-size filters cut into pieces, according to the manufacturer's instructions. DNA was eluted in 50-80 µl of the kit's elution buffer (10 mM Tris, pH8). 18S rDNA fragments of ca. 550 bp, encompassing the V4 hypervariable region, were amplified by PCR using primers EK-565F (5'-GCAGTTAAAAAGCTCGTAGT; Simon *et al.*, in revision) and 18S-EUK-1134-R – UnonMet (5'-TTTAAGTTTCAGCCTTGCG) biased against Metazoa (Bower *et al.* 2004). These primers were tagged with twenty 10 bp Molecular IDentifiers (MIDs) to allow differentiation of PCR products from 20 distinct samples pooled together for multiplex sequencing. PCR were conducted in a total volume of 25 µl using

1.5 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP (PCR Nucleotide Mix, Promega), 0.3 μM of each primer, 0.3 to 3 μl of DNA sample and 0.5 U HotStart Taq polymerase (Taq Platinum, Invitrogen). Twenty-five amplification cycles (94°C for 30 s, 58°C for 45 s and 72°C for 90 s) were preceded by a 3 min denaturation step at 94°C and ended by 10 min at 72°C for final extension. Five to 11 PCR products of each sample were pooled together. Pools were then purified using the QIAquick PCR purification kit (QIAGEN), according to the manufacturer's instructions. Purified amplicons of 20 samples were pooled together at the same amount of 200 to 400 ng depending on the pool. Amplicons were pyrosequenced using the 454 GS-FLX Titanium technology from Roche (Beckman Coulter Genomics).

#### *454-pyrosequence cleaning*

We obtained a total of 1 521 380 reads. To include in our analysis only high-quality sequences, several filters were applied using a local pipeline (Bachy *et al.*, 2013). Pyrosequences from all samples were processed indifferently. First, reads containing errors in primer sequences or positions with undetermined bases were eliminated. The other pyrosequences were processed using AmpliconNoise (Quince *et al.* 2011) integrated in our pipeline to eliminate errors due to both PCR reactions and pyrosequencing. Filtered reads were then clustered into OTUs with 98% similarity cut-off using AmpliconNoise. OTUs composed of only one single sequence were eliminated for precaution. Each OTU was represented by its most abundant sequence, used as reference. Reference sequences were blasted against the PR2 database (Guillou *et al.* 2013) to assign OTUs to taxonomic groups based on sequence similarity. OTUs affiliated to cryptophyte nucleomorph were not included in our analysis. Filtered sequences from all OTUs were then attributed to samples according to their MIDs. Then, potential chimerical OTUs were eliminated using a stringent procedure with both automated and manual steps, previously described in Simon *et al.* (accepted). After filtering, 1 272 748 correct reads were kept (nearly 84% of the initial number of reads).

#### *Statistical analyses*

All statistical analyses were conducted with the R software (<http://cran.r-project.org>) (R Development Core Team 2013). Diversity and richness indices were determined using the “Vegan” R package (Oksanen *et al.* 2013). Richness was evaluated by rarefaction analyses as the estimated number of OTUs in a random subsample of each sequence library, of the same size as the smallest one (Hurlbert 1971). Simpson index was calculated as  $D' = 1 - \sum_{i=1}^S (f_i^2)$  (Simpson 1949) and evenness as  $e = -\sum_{i=1}^S f_i \ln(f_i) / \ln(S)$  (Pielou 1966) with S being the observed number of OTUs and  $f_i$  the frequency of each OTU<sub>i</sub> in the sample. Pairwise Bray-



Curtis dissimilarities between samples were calculated on the basis of OTU percentages (instead of analysing the raw counts of sequences to not consider heterogeneity due to highly different numbers of sequences per sample) or on OTU percentages after a Wisconsin standardization to balance the importance of rare and abundant OTUs (Bray & Curtis 1957). Boxplots were drawn with notches to compare the distribution of Bray-Curtis dissimilarities or diversity and richness estimators between different classes (Chambers *et al.* 1983). Non-metric multidimensional scaling (NMDS) ordination analyses were conducted on Bray-Curtis dissimilarities based on square-root transformed and Wisconsin standardized OTU percentages, using the 'Vegan' R package. Ellipses grouping various classes were drawn using `s.class` from the 'Ade4' R package (Dray & Dufour 2007). Permutational multivariate analyses of variance using distance matrices (Anderson 2001) were conducted using Bray-Curtis dissimilarities calculated on OTU percentages of reads, with the `Adonis` function of the 'Vegan' package. A Venn diagram presenting the number of shared OTUs between ecosystems was drawn using the 'gplots' R package (Bolker *et al.* 2012) on pooled sequences from monthly samples. To check for seasonality, pairwise Bray-Curtis dissimilarities between each sample from the first sampling year and the samples from the same system one to twelve months apart were calculated. Then, the mean and standard errors of Bray-Curtis dissimilarities between samples separated by the same number of months were computed. To compare samples based on their environmental characteristics, Principal component analyses (PCA) were computed using the 'Ade4' R package on centered and scaled values, missing values been replaced by the mean of all values of the same parameter. Nutrient concentrations below the kit detection limit were set as the limit value minus a number chosen randomly in a normal distribution of mean 0 and standard deviation  $10^{-6}$ . Canonical correspondence analysis (CCA) was calculated with 'Ade4' with the environmental parameters after replacement of below detection and missing values and with the Wisconsin standardized percentages of the 37 OTUs that were detected in at least 25% of samples (i.e. 28 samples) and with  $\geq 0.1\%$  mean relative abundance per sample (described in Supplementary Table 2). The proportion of community variance explained by environmental parameters was estimated as the ratio between the sum of all eigenvalues from the CCA and the sum of all eigenvalues of a correspondence analysis (CoA) conducted on the same standardized OTU proportions (Borcard *et al.* 1992).

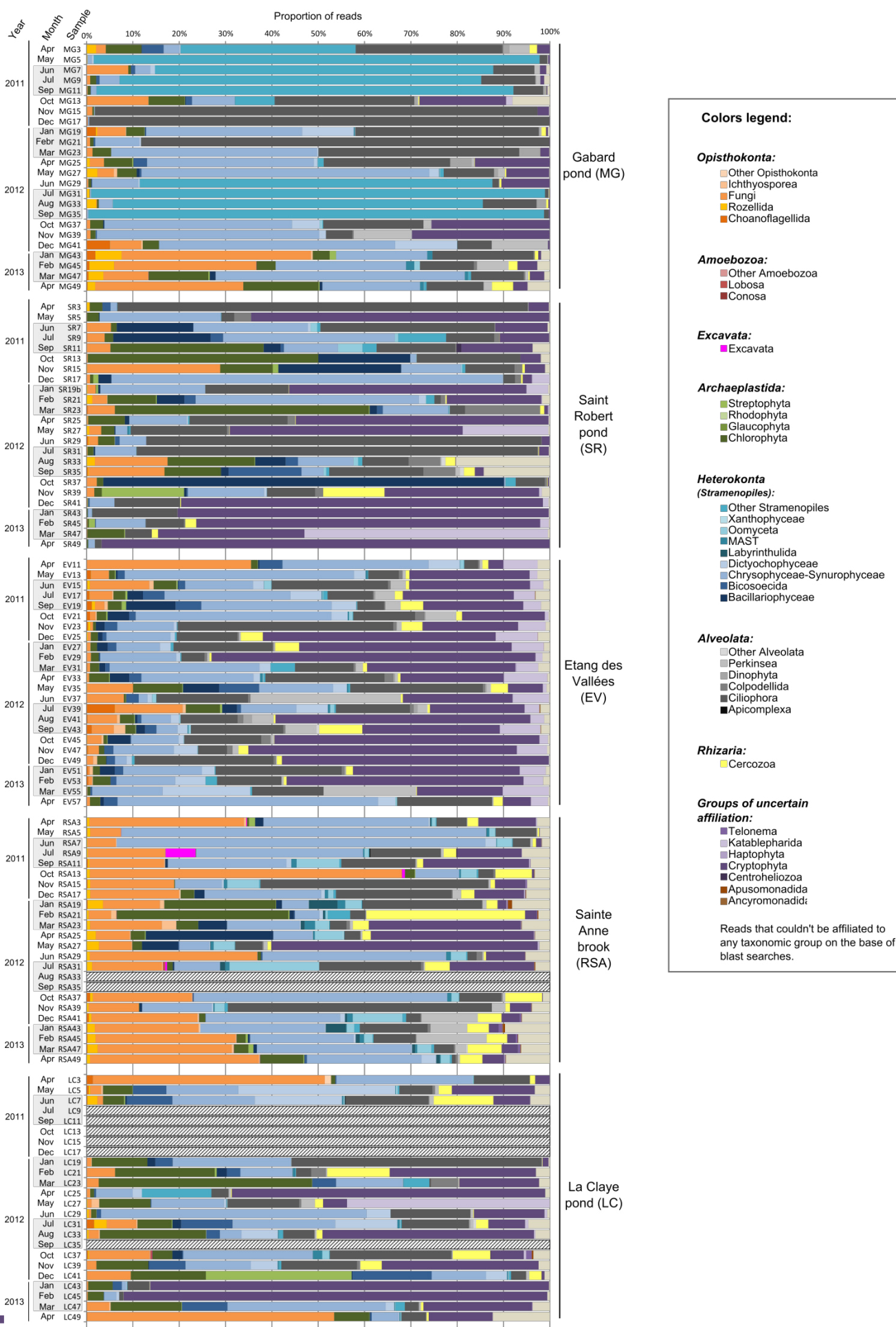
## D- Results

### *Environmental characteristics of the study sites*

Four ponds and one brook were sampled monthly from April 2011 and April 2013. During the sampling period, the studied ecosystems experienced high water temperature variations (Supplementary Figure 1). Each system was characterized by particular environmental conditions. Gabard pond had the lowest pH values (6.2 on average), high chlorophyll *a* and low nitrite concentrations (Supplementary Figures 1, 2A and Supplementary Table 1). Saint Robert pond was characterized by episodic high concentrations of ammonia, reaching up to 2.65 mg l<sup>-1</sup> in July 2012 and January 2013 (Supplementary Figure 1) and very high chlorophyll *a* concentrations at summer and autumn (Supplementary Figure 1). Etang des Vallées was usually the most oxygenated and had the lowest dissolved orthophosphate concentrations (Supplementary Figures 1, 2A and Supplementary Table 1). Sainte Anne brook, which was dry in August and September 2012, had low chlorophyll *a* concentrations and very high conductivity (Supplementary Table 1). La Claye pond was dry from the end of July 2011 to December 2011 and in September 2012 and was characterized by low concentrations of nitrite and ammonia.

**Figure 1 (next page):** Histograms showing the relative abundance of 18S rDNA amplicon reads assigned to high-rank taxa, in all samples from the five ecosystems over the two-years survey. Sainte Anne brook was dried in August and September 2012, and La Claye pond from August to December 2011 and in September 2012; Hatched bars correspond to missing data because of the drought periods: Sainte Anne brook was dried in August and September 2012, and La Claye pond from August to December 2011 and in September 2012.

# Chapitre 3



*Overall composition of the small protists*

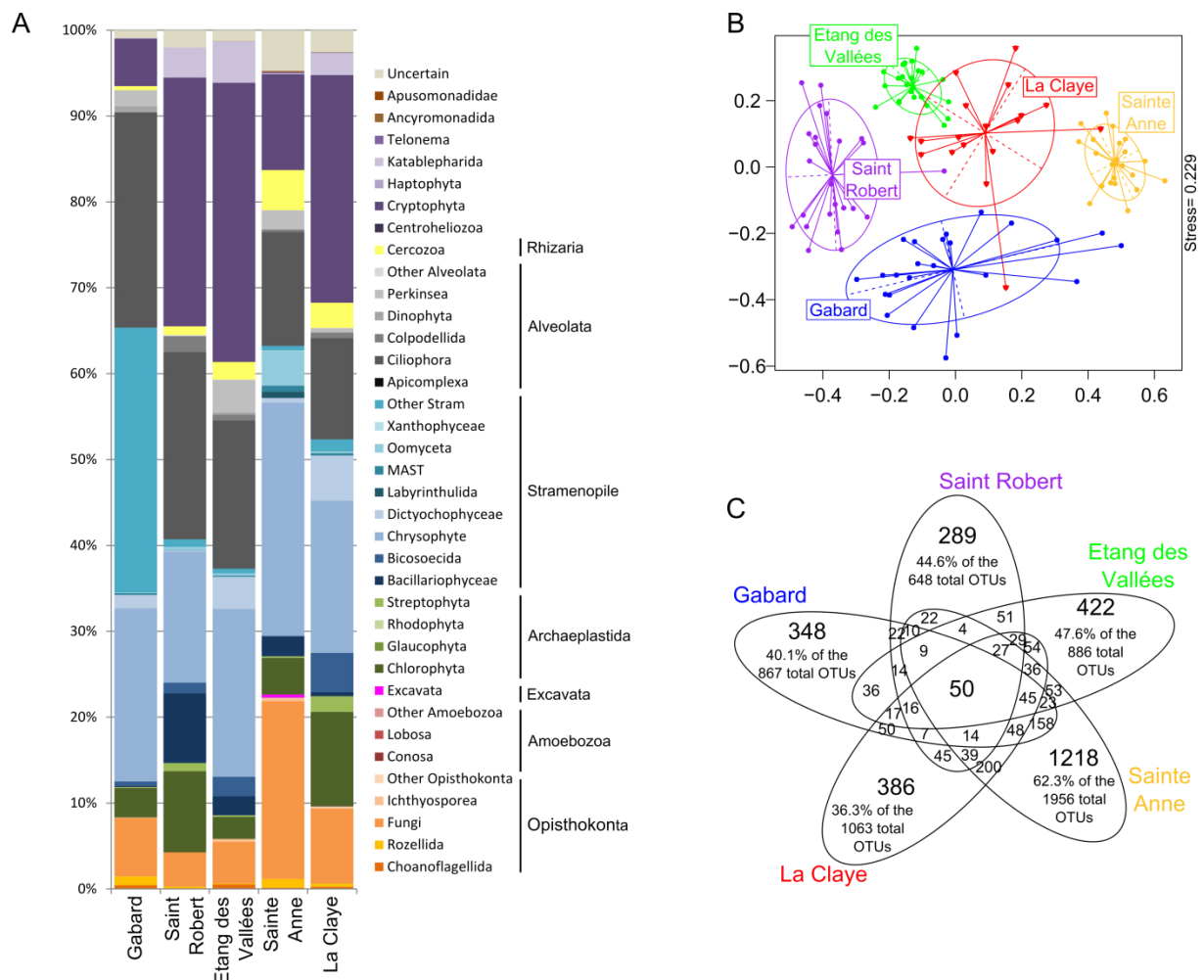
The composition of the small eukaryotes (0.2-5  $\mu\text{m}$  in size) was investigated in the 112 surface water samples collected over the 2-years survey in the five ecosystems, using PCR amplification and 454-pyrosequencing of 18S rDNA fragments. We obtained a total of 1 272 748 quality filtered reads that grouped within 3 742 OTUs (98% identity cut-off). Nearly 93.1% of the OTUs were at low abundance ( $< 0.05\%$  of the reads per sample, on average) (Supplementary Figure 3A) when only 0.7% of the OTUs (26 OTUs) were abundant ( $> 0.5\%$  of the reads per sample) (Supplementary Figure 3B) and represented together 51.4% of the reads per sample, on average. As little as 10.6% of the OTUs could not be assigned to a taxonomic group based on sequence similarity, representing 2.3% of the reads per sample, on average.

Detected OTUs were affiliated to all currently recognized supergroups (Adl *et al.* 2012): Opisthokonta, Amoebozoa, Excavata, Archaeplastida, SAR (Stramenopiles-Alveolata-Rhizaria), along with group of uncertain position in the eukaryotic tree (Cryptophyta, Centroheliozoa, Telonema, Haptophyta, Katablepharida, Ancyromonadida and Apusozoa). Stramenopiles, alveolates, cryptophytes, opisthokonts, and archaeplastids were the most abundant groups during the survey. Stramenopiles represented as much as 35.1% of the reads per sample, on average, and were highly diverse with 956 OTUs affiliated to this supergroup (25.5% of all OTUs). Most of these OTUs were affiliated to Chrysophyceae-Synurophyceae, the most abundant stramenopile group (20.0% of the reads per sample, on average). Bacillariophyceae, Raphidophyceae, Dictyochophyceae, Bicosoecida and Oomyceta also represented more than 1% of the reads per samples, on average (Supplementary Figure 3C). Less abundant OTUs were also affiliated to other groups such as MAST lineages, Labyrinthulida or Xanthophyceae. Alveolates was the second most abundant supergroup, representing 20.9% of the reads per sample, on average. They were represented by several phyla (i.e. Perkinsea, Colpodellida, Dinophyta, and Apicomplexa) (Figures 1 and Supplementary Figure 3C), but were dominated by ciliates both in terms of abundance (18.2% of the reads per sample, on average) and diversity (10.8% of all OTUs). The parasitic Perkinsea was occasionally highly abundant, reaching up to 32.4% of the reads in Etang des Vallées in June 2012 (Figure 1). Cryptophytes were as abundant as Alveolates (representing 20.8% of the reads per sample, on average) (Supplementary Figure 3C) but were not highly diverse, with only 1.4% of the OTUs affiliated to this group. Also, opisthokonts represented 9.9% of the reads per sample, on average, and were largely dominated by fungi (8.8% of the reads per sample, on average) (Supplementary Figure 3C), especially chytridiomycetes, ascomycetes and basidiomycetes. Rozellids, choanoflagellates, and ichthyosporeans were less

abundant (mean relative abundance per sample  $< 0.5\%$ ), but could occasionally reach higher proportions, such as the Rozellids (5.7% of the reads in Gabard pond in January 2013) or the choanoflagellates (6.1% of the reads in Etang des Vallées in July 2012) (Figure 1). Archaeplastida, especially chlorophytes, were detected at high proportions (Supplementary Figure 3C). Several OTUs of streptophytes, rhodophytes and glaucophytes, were also occasionally recorded (Figure 1). Glaucophytes were represented by OTU 2122 only, affiliated to *Cyanophora paradoxa* with 99.6% of sequence identity, and detected twice at low relative proportion in Saint Robert pond. Rhizarians and katablepharids were detected at a mean proportion per sample of ca. 2% (Supplementary Figure 3C). Katablepharids were always relatively abundant in Etang des Vallées (4.8% of the reads, on average) (Figure 1) and were occasionally at extremely high proportions elsewhere (52.5% of the reads in Saint Robert pond in March 2013 for example) (Figure 1). Rhizarians were only represented by the fairly diverse cercozoans, with 7.7% of the detected OTUs affiliated to this group. Excavata were rather rare in general (Supplementary Figure 3C), at the exception of OTU 1147, affiliated to *Trimastix pyriformis* (98.6% of sequence identity), which represented 6.6% of the reads in Sainte Anne brook in July 2011. Amoebozoan, haptophytes, centroheliozoans, telonemids, apusomonads and ancyromonads were also detected though at lower diversity and abundances (Supplementary Figure 3C). Centroheliozoan could only be detected at one sampling date in Saint Robert and La Claye ponds, and twice in Sainte Anne brook. Similarly, OTUs affiliated to Conosa, Excavata, Rhodophyta and Apicomplexa were rarely detected (in less than 10 samples).

#### *Contrasted communities in the different ecosystems*

The small eukaryote communities differed in the 5 ecosystems as revealed by clustering of the samples by ecosystems in a NMDS analysis based on Wisconsin standardized OTU proportions (Figure 2B). More than 20% of the total variation between samples was due to the variance between ecosystems, as estimated by a permutational multivariate analysis of variance using a Bray-Curtis distance Matrix (Table 1). Etang des Vallées and La Claye pond seemed to host the most similar communities (Figure 2B), harbouring the denser cryptophyte communities along with Saint Robert pond. Gabard pond was dominated by stramenopiles (especially Raphidophyceae) and ciliates, and Sainte Anne brook by Chrysophyceae-Synurophyceae and fungi (Figure 2A).



**Figure 2: Distinct communities in the 5 ecosystems.** A: Taxonomic compositions of the 5 ecosystems. All samples from the 2-years survey are pooled for each system, after correction for different numbers of reads in different samples. B- Nonmetric Multi Dimensional Scaling plot, built on square-root transformed and Wisconsin standardized Bray-Curtis distances between all samples. C: Venn diagram showing the number of OTUs shared by several ecosystems or specific to an ecosystem. Proportions of OTUs specific to each ecosystem among all OTUs detected in this ecosystem are indicated below. All samples from the 2-years survey are pooled for each ecosystem.

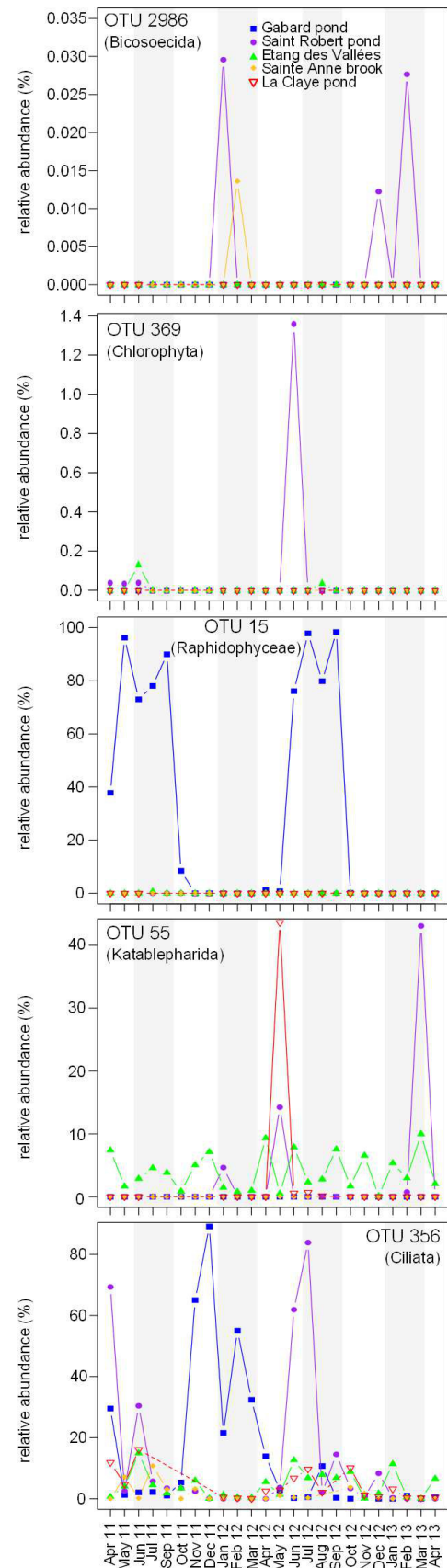
Fifty OTUs only, out of the 3 742 OTUs recorded during the 2-years survey, were detected in the five ecosystems (Figure 2C). They were affiliated to opisthokonts (mostly fungi), archaeplastids (mostly chlorophytes), stramenopiles, alveolates (mostly ciliates), cercozoans (OTU 20 only), cryptophytes and katablepharids (OTU 55 only). They were neither continuously abundant nor even present at the same time in all ecosystems, as OTU 356 for example, affiliated to ciliates (Figure 3). As much as 2 663 OTUs (more than 71% of all the OTUs) were specific to one ecosystem (Figure 2C), out of which 1 180 OTUs (31.5% of all OTUs) were detected at least twice. Especially, 62.3% of the OTUs retrieved in Sainte



### Chapitre 3

Anne brook were not detected elsewhere. Also, some OTUs reached very high abundances in one system without being detected in their closest neighbour, as exemplified in Gabard pond by OTU 15, affiliated to the Raphidophyceae (Figure 3). Differences between ecosystems seemed to be due to the less frequent OTUs more than to the most abundant OTUs. Indeed, when the percentages of reads for the different OTUs were not Wisconsin standardized before conducting the NMDS (i.e. the more an OTU is abundant in at least one sample, the more it has an impact on the comparison between samples), samples were no longer clustered per ecosystems (data not shown). In addition, abundant OTUs ( $> 0.5\%$  of the reads per sample, on average) were detected in more ecosystems (median of 4.5, min. of 1 and max. of 5) than less frequent OTUs ( $< 0.5\%$  of the reads per sample, on average; median of 1, min. of 1 and max. of 5).

**Figure 3 (on the right):** Relative abundance of five OTUs showing different types of dynamics, in samples from the five ecosystems, over the two-year survey. Each colour corresponds to the relative abundance (set as the proportion of reads) dynamics in one of the sampled systems (see legend in the first figure). Taxonomic affiliations of OTUs are indicated between brackets. Sainte Anne brook was dried in August and September 2012, and La Claye pond from August to December 2011 and in September 2012. Missing data are indicated by dashed lines.



The small eukaryote communities of the five ecosystems differed also in richness and diversity. Richness was much higher in the brook than in any other ecosystems

(Supplementary Figure 4 and 5) while Gabard and Saint Robert ponds had the lowest richness (of 50.3 and 38.0 OTUs per sample, on average, respectively). La Claye pond and, especially Etang des Vallées and Sainte Anne brook, were not only richer but also more diverse than Gabard and Saint Robert ponds (Supplementary Figure 4).

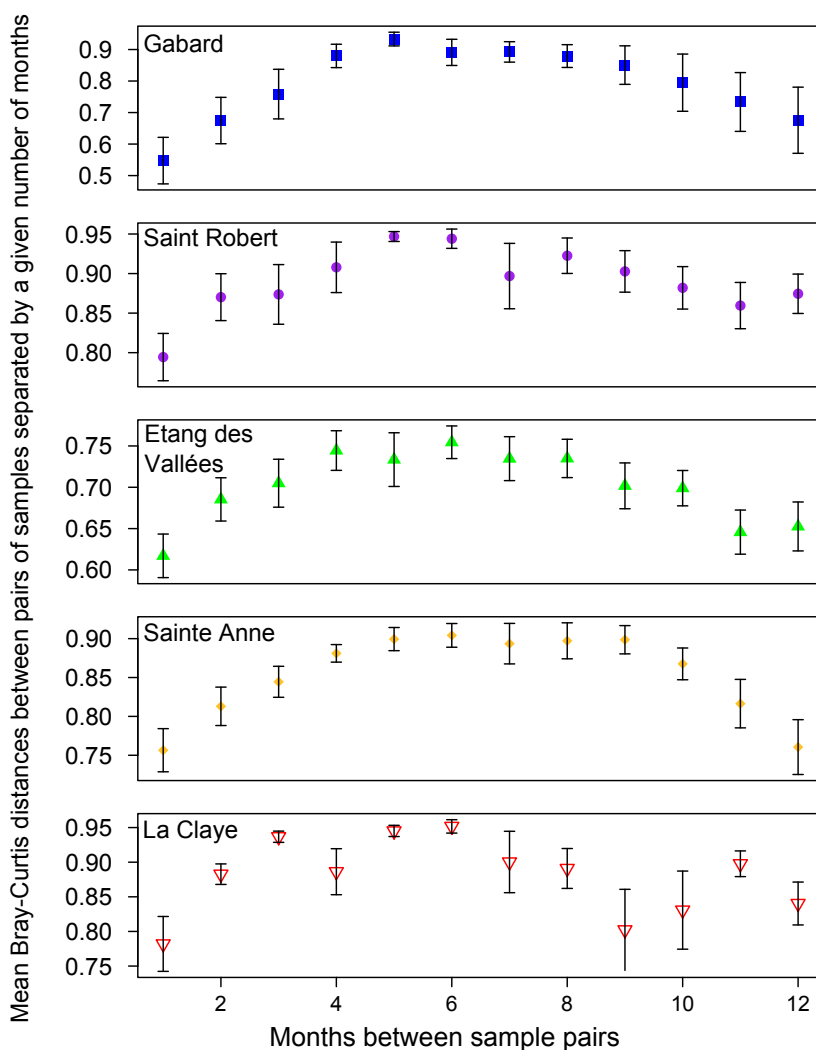
Overall, if the small eukaryote communities allowed differentiating the ecosystems, the measured environmental as a whole did not. Gabard pond was the only system that was slightly different from the other systems according to a Principal Correspondance (PCA), being characterized by low nitrite concentration, pH and conductivity (Supplementary Figure 2A).

#### *Highly dynamic small eukaryote communities*

Beyond the differences between ecosystems, the small eukaryote communities of all the ecosystems studied were highly dynamic. The mean Bray-Curtis dissimilarities between all pairs of samples per ecosystem were high, ranging from 0.71 for Etang des Vallées to 0.89 for Saint Robert pond (Supplementary Figure 6A). The dissimilarities reached higher values after a Wisconsin standardization of the OTU proportions, of 0.88 for Etang des Vallées to 0.96 for Gabard and Saint Robert ponds, suggesting that the estimated relative abundance of low frequent OTUs varied more than the one of frequent OTUs (Supplementary Figure 6B). In both computations of the dissimilarities, the small eukaryote community seemed to be more stable in Etang des Vallées than in the other systems, since Bray-Curtis distances between all pairs of samples from that pond were lower than from each other system (Supplementary Figure 6). Nevertheless, the dynamics of the eukaryote communities were rapid and drastic in all systems, the median of pairwise Bray-Curtis dissimilarities between successive samples varying from 0.61 for Etang des Vallées to 0.79 for Saint Robert and La Claye. Previously undetected OTUs were observed each month (Supplementary figure S5). Also, the community structure was very variable over the 2 years in all systems as revealed by the strong variations in richness, varying of more than 10 times in Gabard pond (Supplementary Figure 4).

The small eukaryote communities did not differ markedly between the two years of the survey in each system, with a maximum of 9.8% of the total variance between samples corresponding to differences between the two sampling years (recorded in Gabard pond, Table 1). A seasonal pattern of the whole small eukaryote community composition was revealed for each system by pairwise Bray-Curtis dissimilarities. Indeed, the dissimilarity values were the highest between samples separated by 5 to 6 months (i.e. samples from opposite seasons) and the lowest between samples close to each other or separated by 12

months (i.e. samples from the same season, Figure 4). This pattern was less smooth for La Claye pond likely because of the lack of data during the 6 months of drought. The strength of the seasonality differed between the ecosystems but was always high, differences between seasons corresponding to 19.2% for Saint Robert to 35.8% for Gabard pond of the variation in the small eukaryote communities for each ecosystem (Table 1). The small eukaryote community differed more between summer and winter than between autumn and spring, as revealed by an NMDS analysis realized for each ecosystem (Supplementary Figure 7). The environmental conditions in each ecosystem were also characterized by seasonality as disclosed by PCA (Supplementary Figure 2, B to F), suggesting that the temporal dynamics of the small eukaryotes could result from the seasonal variations of the physico-chemical parameters measured.

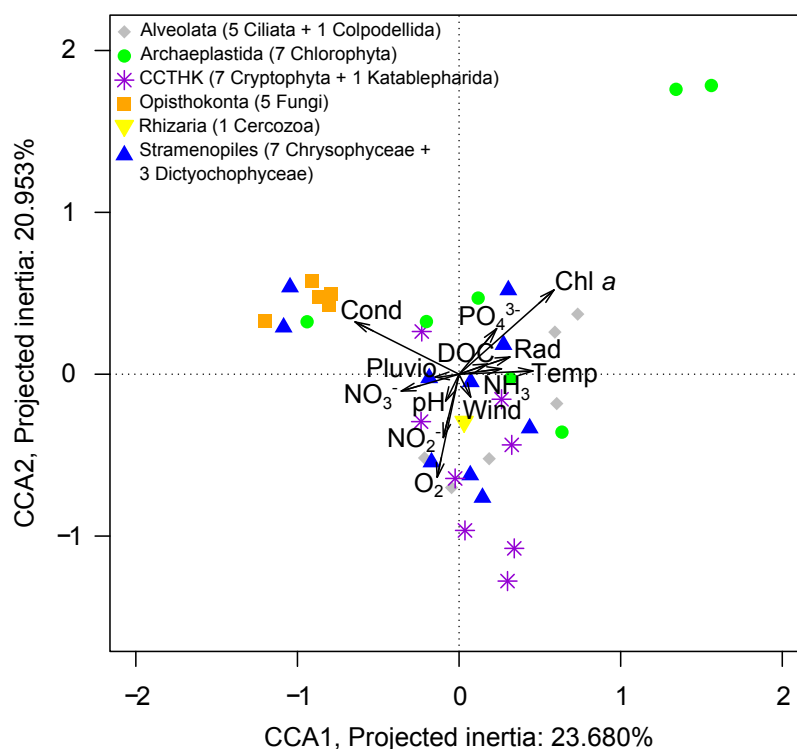


**Figure 4: Pairwise Bray-Curtis dissimilarities between samples separated from 1 to 12 months.** Each point represents the mean of pairwise Bray-Curtis dissimilarities between samples one to twelve months apart, plotted with standard errors. All means are based on 12 dissimilarities except for Sainte Anne brook and La Claye pond because of missing data during the drought period.

#### *Rapid successions of OTUs, taxonomic and functional groups*

The relative proportions of diverse taxonomic groups fluctuated greatly and rapidly over the 2-years survey (Figure 1 and Supplementary Figure 8), often supported by a handful of OTUs only, varying occasionally of more than a 100 times in two consecutive months (e.g.

cryptophytes in La Claye pond in December 2012 and January 2013, Supplementary Figure 8). Globally, taxonomic groups were rarely detected at high relative proportions in more than two consecutive months. Successive peaks of relative abundances were either due to the rapid dynamics of the same OTUs (e.g. katablepharids in Etang des Vallées) or to raises in proportions of distinct OTUs (e.g. bicosoecids in Saint Robert pond). In addition, groups that were usually low frequent could reach discrete high relative abundances. This was the case of the excavates in Sainte Anne brook in September 2011 (Figure 1 and Supplementary Figure 8), due to OTU 1147.



**Figure 5: Canonical Correspondence Analysis (CCA) plot.** CCA was conducted on all samples and on the 37 OTUs detected in 25% of samples and having a mean relative abundance per sample of 0.1% or above, after a Wisconsin standardization. Both physico-chemical parameters measured in the ecosystems' water and weather information were included in the analysis. Dots represent OTUs, with colour and form indicating their taxonomic affiliation. CCTHK: Cryptophyta, Centroheliozoa, Telonema, Haptophyta, Katablepharida. Chl a: Chlorophyll a, DOC: Dissolved Organic Carbon, Cond: Conductivity, Pluvio: Pluviometry during the week before sampling, Wind: Mean wind speed at 10m high during the sampling day, Rad: Mean daily solar radiation, Temp: water temperature.

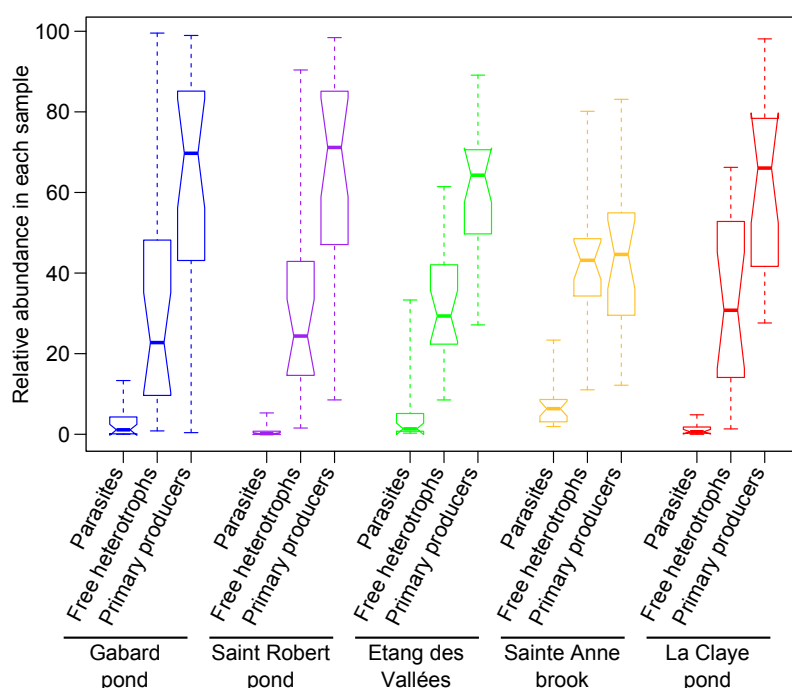
To decipher the potential role of the environmental conditions on the dynamics of the small eukaryotes in small freshwater ecosystems, we measured a panel of physico-chemical parameters at surface and used meteorological data recorded at the closest meteorological stations. We investigated the relationships between all these environmental parameters and the dynamics of the 37 widespread OTUs which represented at least 0.1% of the reads per

sample, on average, and that occurred in at least a quarter of the samples (Supplementary Table 2). Physico-chemical data only and both physico-chemical plus climatologic parameters explained 18.7% and 21.9% of the variance of these OTUs, respectively. More precisely, if fungi were strongly associated to high conductivity, the opposite pattern characterized the ciliates, as revealed by a CCA (Figure 5). Also, the proportions of chlorophytes seemed to correlate with high concentrations of chlorophyll *a* and phosphate. Cryptophytes were mostly abundant when the dissolved dioxygen and nitrite concentrations were higher, but their association with environmental variables was less clear.

At a finer scale, OTUs were also showing a high dynamics. Indeed, only a few OTUs were constantly recorded during the survey: 1, 0, 2, 1 and 7 in Gabard, Saint Robert, Sainte Anne, La Claye and Etang des Vallées (e.g. OTUs 55 and 356, Figure 3), respectively. In addition, the OTUs had very diverse dynamics. Within the low frequent OTUs, two types of dynamics were observed. First, 275 OTUs, representing 7.3% of all the OTUs, always represented less than 0.05% of reads per sample, although they were detected in at least two samples of the whole survey (e.g. OTU 2986, affiliated to stramenopiles, Figure 3). Around a quarter of these OTUs were affiliated to fungi, while others were spread among all supergroups. Secondly, several OTUs that were at low abundances most of the time could occasionally reach high relative proportions (e.g. OTU 369, Figure 3). Especially, 47 OTUs (1.3% of all the OTUs) represented more than 0.5% of the reads in one sample and were not detected elsewhere. These OTUs were mainly affiliated to fungi and stramenopiles as well as to cercozoans, chlorophytes, Perkinsea, excavates and centroheliozoans. Then, among the most abundant and frequent OTUs, some presented a marked seasonality. In particular, OTU 15, affiliated to the flagellate algae *Gonyostomum semen* (Raphidophyceae), bloomed in Gabard pond during both summers (Figure 3) at very high proportions (up to 98% of the reads in September 2012). In addition, a few OTUs remained abundant during the 2-year survey and without seasonal pattern: 3, 4, 6, 5 and 15 OTUs represented more than 0.5% of the reads in at least half of the samples in Gabard, Saint Robert, Sainte Anne brook, La Claye and Etang des Vallées (e.g. OTU 55, Figure 3), respectively. These OTUs were affiliated to chrysophytes and fungi in Sainte Anne brook, and mainly to ciliates, cryptophytes and stramenopiles elsewhere. Interestingly, a small set of OTUs that were always detected and abundant in some systems were low frequent or episodic in other environments as for the OTU 55, affiliated to the katablepharids (Figure 3).

Finally, to get an insight of the ecology of our five small systems, the OTUs were attributed to three functional groups on the basis of their taxonomic affiliation: putative parasites, other heterotrophs and primary producers. Over the two years, primary producers

were more abundant than heterotrophs in all systems but Sainte Anne brook (Figure 6). The relative proportion of these functional groups was highly dynamic over time (Supplementary Figure 9). In Gabard pond, primary producers were highly dominating in summer, due to *Goniostomum semen* blooms. Parasites had generally low relative proportions in Saint Robert, La Claye and Gabard ponds, at the exception of November 2012 to February 2013 in Gabard pond where Perkinsea and Rozellida were recorded at significant proportions. Parasites, perkinsea especially, reached important proportions in Etang des Vallées, and Oomycetes, Perkinsea and Rozellids, mainly, were non-negligible components of communities in Sainte Anne brook most of 2 years.



**Figure 6: Distribution of relative abundances of potential parasites, other heterotrophs and primary producers in all samples from the five ecosystems.** The thickest line inside each box represents the median of the distribution, bottom and top borders of boxes correspond to the first and third quartiles and whiskers extend to minimal and maximal distances. Notches are drawn to indicate whether medians of the three distributions can be considered as different for each ecosystem (Chambers et al., 1983, p.62). Putative parasites: Apicomplexa, Rozellida, Oomyceta, Perkinsea, Ichthyosporea ; potential other heterotrophs: Ancyromonadida, Apusomonadida, Bicosoecida, Centrohelioczoa, Cercozoa, Choanoflagellida, Ciliophora, Colpodellida, Conosa, Discosea, Fungi, Katablepharida, Labyrinthulida, Lobosa, Malawimonadida, MAST, Metamonada, Metazoa, Telonema ; putative primary producers: Bacillariophyceae, Chlorophyta, Chrysophyceae-Synurophyceae, Cryptophyta, Dictyochophyceae, Dinophyta, Glaucophyta, Haptophyta, Rhodophyta, Streptophyta, Xanthophyceae.



## E- Discussion

*Wide small-eukaryote diversity in shallow freshwater ecosystems highlighted by temporal survey*

A few previous snapshot studies conducted in ponds revealed the presence of a wide diversity of the small-eukaryotes, including lineages of particular interest (Šlapeta *et al.*, 2005; Simon *et al.*, 2013, Simon *et al.*, accepted). OTUs detected in the five studied ecosystems were affiliated to all accepted supergroups (Adl *et al.* 2012) including infrequent taxa such as glaucophytes (Richards *et al.* 2005; Massana 2011; Triadó-Margarit & Casamayor 2012). The three most abundant groups of all our shallow environments were stramenopiles, alveolates and cryptophytes, which are abundant members of eukaryotic assemblages in freshwater lakes at all seasons (Lefranc *et al.* 2005; Richards *et al.* 2005; Šlapeta *et al.* 2005; Triadó-Margarit & Casamayor 2012). Fungi are also a common and important functional group in freshwaters (Lepère *et al.* 2008; Mangot *et al.* 2012; Taib *et al.* 2013). We detected them at relatively high proportions (Supplementary Figure 3C), especially in Sainte Anne brook, a really shallow and narrow stream (Figure 1 and 2A). These especially high proportions in the brook could be explained by the importance of the littoral and benthic zones that may provide high content of organic matter, or result from a contamination by spore-containing sediment. Chlorophytes were also important in our shallow systems as they are in marine and freshwater environments (Johnson & Sieburth 1982; Stockner 1988; Massana 2011). Stramenopiles, alveolates and chlorophytes are also widely distributed in marine habitats (Vaulot & Eikrem 2008; Kirkham *et al.* 2013), while cryptophytes and fungi especially are usually found less abundant (Massana 2011) at least in open ocean (Vaulot & Eikrem 2008).

Beyond these few abundant groups, several taxa such as Centroheliozoa, Glaucophyta or Excavata were detected at low abundance and in very few samples (Supplementary figure S8), and the vast majority of the 3 742 detected OTUs occurred at low frequency (less than 0.05% of reads on average per sample). Due to the difficulty to detect very low abundant phylotypes in each snapshot, even using Next Generation Sequencing (NGS) technologies, and because new organisms are potentially continuously colonizing systems from other sites, temporal surveys are expected to allow the detection of an even wider diversity than in previous discrete studies (Nolte *et al.* 2010). Indeed, the number of OTUs detected in one ecosystem during the whole survey was 6 to 11 times higher than the average number of OTUs observed at any discrete date in the same system, and each new monthly sampling allowed the detection of previously unseen phylotypes (Supplementary Figure 5).

Communities differed significantly between the five ecosystems over the 2-year survey (Table 1 and Figure 2B), especially on the basis of low-frequent OTUs. Sainte Anne brook was an exception as its small eukaryote community clustered apart from the pond's communities, even based on data that were not Wisconsin standardized (data not shown). This suggests that both abundant and low frequent OTUs differed in the brook and in the ponds. Studies of other ecosystems will be needed to know whether it is a real difference between the two kinds of environments or if Sainte Anne is a particular case. Fifty OTUs only were shared by the five systems and a high number of OTUs were specific to one system only, with as many as 1 218 OTUs detected only in Ru Sainte-Anne for example (Figure 2C). The small-eukaryote community composition did not differ among the ecosystems depending on their geographical distances, Etang des Vallées and La Claye ponds having close community composition (Figure 2 C) while they are some of the most distant systems (9km apart). Small and shallow freshwater ecosystems are characterized by an important physico-chemical range (Céréghino *et al.* 2008). Here, the measured environmental properties as a whole did not allow differentiating the 5 ecosystems (Supplementary Table 1, Supplementary Figure 1 and 2A), suggesting that other parameters are needed to better explain the large differences in communities between sites.

#### *Hectic and seasonal dynamics of both abundant and low frequent small eukaryotes*

We first observed drastic and rapid changes in the small eukaryote community composition of all our systems, as revealed by the high dissimilarity values between consecutive samples. Such rapid changes have also been recorded in oceans (Vigil *et al.* 2009) and lakes (Lepère *et al.* 2006), even at a few days scale (Mangot *et al.* 2012). These community composition shifts might be driven by the great and rapid variations of the measured local environmental parameters (Supplementary Figure 1) as well as potential biotic factors such as predation, parasitism, viral lysis and exudates of other microbes or metazoans.

Deciphering seasonal microorganism dynamics is a tedious task that requires an important sampling and analysis effort. To date, no consensus has been found on microbial seasonality, for both prokaryotes and eukaryotes, in marine and freshwater environments, whatever the molecular methods (i.e. fingerprinting, cloning-sequencing, NGS). Although microbial communities sometimes show seasonal dynamics in oceans and lakes (Gilbert *et al.* 2012; Jones *et al.* 2013; Kim *et al.* 2014), putative seasonality does not necessarily repeat over consecutive years (Lepère *et al.* 2006) and inter-annual differences can be more important than differences between seasons (Kim *et al.* 2014). All our ecosystems were greatly influenced by seasons, as shown by the surface water temperature dynamics, and

winter and summer seemed to be characterised by opposite environmental features (Supplementary Figure 2 and Table 1). Taken individually, the physico-chemical parameters were not following any marked seasonal pattern, at the exception of nitrate and nitrite concentrations. Strikingly, the small eukaryote community composition was relatively repeatable over the 2-years survey and was significantly varying with seasons, especially with the two opposite seasons winter and summer (Table 1, Figure 4 and Supplementary Figure 7). Moreover, the small eukaryote community was more structured by seasons than by year in all ecosystems, at the exception of Saint Robert pond (Table 1). Then, a seasonal dynamics seemed to superimpose over the rapid hectic dynamics also observed. Seasonal parameters, maybe temperature and/or irradiance, appear to be more important than occasional events in most of our small and shallow systems during the sampling period. However, we did not observe a general seasonal pattern for the richness in our systems as observed in others surveys, for bacteria or eukaryotes (Gilbert *et al.* 2012; Jones *et al.* 2013), but peaks in richness observed at all seasons instead (data not shown).

Effects		All	Gabard pond	Saint Robert pond	Etang des Vallées	Sainte Anne brook	La Claye pond
Samples	F.Model	8.3601					
	R2	0.20631					
Ecosystems	F.Model	2.5939	4.4853	1.6863	2.19928	1.9582	1.5829
	R2	0.04738	0.35753	0.19161	0.24628	0.23905	0.24404
Seasons	F.Model	2.0872	3.773	2.4289	1.2112	1.2828	1.1955
	R2	0.01271	0.09807	0.092	0.04521	0.0522	0.06658
Sampling years	F.Model	2.1275					
	R2	0.15545					
Ecosystems x seasons	F.Model	1.9274					
	R2	0.04694					
Ecosystems x sampling years	F.Model	1.2306	1.6486	0.9714	0.99371	1.1392	1.2073
	R2	0.02246	0.41586	0.11038	0.11128	0.13907	0.12408
Sampling years x seasons	F.Model						
	R2						

$$^1\text{F.Model} = \frac{\text{Sum of squares of distances between classes} / \text{degrees of freedom}}{\text{Sum of squares of distances inside classes} / \text{degrees of freedom}}$$

$$^2\text{R2} = 1 - \frac{\text{Sum of squares of distances inside classes}}{\text{Sum of squares of distances between all sample pairs}}$$

Interactions between effects were estimated and are indicated by “effect A x effect B”

**Table 1: Variance partitioning between effects (Adonis)**

Seasonality is known for various small eukaryote taxa (e.g. Bamforth 1958; Hitchman and Jones 2000; Lara *et al.* 2011), but succeeding to unveil such individual ecological trait in complex environmental microbial communities remains a difficult exercise. Despite our large dataset obtained thanks to an important sampling effort, we were unable to observe a marked seasonality for most of the high-rank taxonomic groups we looked at (Figure 1 and Supplementary Figure 8). A few groups, such as the choanoflagellates, labyrinthulids or dictyochophytes, had a seasonal-like dynamics, but it was not reproducible in several ecosystems and the peaks were sustained by only one or a small set of OTUs. This lack of

pattern in the temporal dynamics may result from the long evolution history of these high-rank taxonomic groups that may comprise taxa with diverse ecological preferences, but also from the local conditions that differ from one ecosystem to another. At a finer scale, we detected a clear seasonality for a few OTUs that were affiliated to various phylogenetic groups while some OTUs stayed at relatively constant proportions. For example, OTU 15, affiliated to the mixotrophic algae *Gonyostomum semen* (Raphidophyceae) was highly seasonal, dominating Gabard pond small-eukaryote community both summers. *G. semen* is known for its preference for low pH and coloured waters (Cronberg *et al.* 1988; Trigel *et al.* 2013), which characterises Gabard pond and may explain why we only detected it 5 times at low frequencies (<0.6% of the reads) in other systems and never in Saint Robert pond, the closest ecosystem. Also, *G. semen* was detected from very geographically distant and ecologically similar systems (Cronberg *et al.* 1988; Cronberg 2005) and our OTU 15 have a high sequence similarity with the strain NIES-1360, isolated in Japan. All these observations are in strong agreement with the Baas-Becking hypothesis “*Everything is everywhere, but, the environment selects*”.

Communities detected in our five systems presented rapid dynamics and seasonal variations, likely due to the superimposition of local parameters and climatic factors effects. These high dynamics could affect all phylotypes irrespectively of their relative abundance (Figure 3, Supplementary figure 6). As previously observed (Nolte *et al.* 2010), temporal variability of the observed community compositions was not only due to changes in abundance of frequently detected OTUs but also to the appearance and disappearance of many phylotypes from the pool of detected organisms (Figure 3). It is impossible with our dataset to assess whether the constant observation of previously undetected OTUs (Supplementary figure 5) is due to immigrating taxa that eventually developed (Pedrós-Alió 2006) or to low frequent small eukaryotes that are constantly present but only occasionally observed. Some OTUs seemed to correspond to this category, being detected several times all along the survey (e.g. the extreme case of OTU 15 Figure 3). In this case, they may always be present in abundances close to the detection limit and be detected a few times by chance, or occasionally emerge because of favourable environmental conditions. Jones and McMahon (2009) estimated that the contribution of bacteria immigrating *via* atmospheric transport to the bacterial community dynamics in lakes was not significant. However, microbes, among which small eukaryotes, can also disperse by other ways. Interestingly, the systems where we detected the highest number of new OTUs each month were La Claye pond, closely surrounded by other ponds from which new organisms could immigrate, and, especially Sainte Anne brook. In the latter, communities at the sampling site are strongly connected to

upstream locations with new organisms arriving continuously thanks to the water flow. Also, we cannot exclude that some small eukaryotes are transported from the surrounding soil because of the erosion caused by the water flow but also by the animals coming especially in this system for their water needs. So, the impact of immigration on small-eukaryote community dynamics might be non-negligible.

#### *Ecology of small-eukaryote assemblages in small and shallow freshwater ecosystems*

Despite the important shade in the forest, small eukaryotes from the ponds were dominated by OTUs affiliated to putative photosynthetic taxa (Figure 6). On the contrary, small-eukaryote communities from lakes and marine systems appeared dominated by heterotrophic cells (Lefranc *et al.* 2005; Lepère *et al.* 2006; Vaulot & Eikrem 2008). Primary producers comprised obligate phototrophs (e.g. chlorophytes) and mixotrophic organisms (e.g. cryptophytes). Putative heterotrophs were affiliated to taxa feeding by phagotrophy (i.e. ciliates, choanoflagellates) or osmotrophy (i.e. most fungi) or to parasite groups. Putative parasites were recorded in all systems, as it has been reported for large lakes. They were affiliated to groups well-known in lakes such as the Perkinsea (Mangot *et al.* 2010), LKM11-Rozellida (e.g. Lepère *et al.* 2008; Taib *et al.* 2013), Ichthyosporea (e.g. Triadó-Margarit *et al.*, 2012), Oomycota (e.g. Richards *et al.*, 2005) and Apicomplexa (e.g. Triadó-Margarit *et al.*, 2012). Small and shallow ecosystems might thus have an important role in the C-cycle because of CO<sub>2</sub> fixation through the microbial eukaryotean important photosynthetic activity (although other members of pond communities such as bacteria will have to be considered). That potential role could be amplified by the higher rate of organic carbon sequestration observed in small than in larger lakes (Downing 2010).

To assess ecological traits of small eukaryote taxa, we selected the 37 OTUs (Supplementary Table 2) that were representing at least 0.1% of the reads (average value per sample) and that were detected in at least a quarter of samples in order to have enough usable information for each OTU and enough OTUs to identify ecological preferences. Although some OTUs had very different environmental preferences, even within the same high-rank taxonomic group, it was possible to discern a few correlations between some of these groups and several environmental parameters (Figure 5). The clearest association was observed between high conductivity values and high proportions of Fungi, which could result from their osmotrophic nutrition or from the contamination of the samples with sediments. Also, cryptophytes seemed to occur in the most oxygenated waters, as previously observed in ponds of different oxic levels (Šlapeta *et al.* 2005). The detection of only few clear correlations for the small eukaryotes in our ecosystems may result from the lack of other parameters, such as

the interactions with other microbes (both prokaryotes and eukaryotes) or metazoans, including predation, mutualism, parasitism (Langenheder & Jürgens 2001; Kühn *et al.* 2004; Foster *et al.* 2006). Indeed, our set of environmental parameters explained only a fifth (22%) of the variation of our 37 OTUs.

Since the emergence of high-throughput sequencing technologies, low frequent taxa, sometimes called rare taxa, have attracted an increasing attention. Indeed, the vast majority of the microbial communities in all kind of environments are composed of these low frequent taxa (Sogin *et al.* 2006; Stoeck *et al.* 2009; Caron & Countway 2009b; Mangot *et al.* 2012; Lepère *et al.* 2013). Because of the important biases generated by the NGS technologies (Gilles *et al.* 2011) that lead to a significant overestimation of low frequent phylotypes (Kunin *et al.* 2010; Massana 2011), we applied a very stringent selection of reads to eliminate most of the artefactual OTUs. The small eukaryotes in our shallow system were still dominated by low-frequent phylotypes (93.1% of OTUs represented <0.05% of reads per sample on average, Supplementary Figure 3A). It is impossible with our dataset to distinguish OTUs corresponding to real components of small-eukaryote assemblages of our ecosystems or to free DNA / dead cells, from the same ecosystem or from other sites, especially for the phylotypes detected only once. However, within these low frequent OTUs, some remained at low frequency (<0.05% of the reads) but were detected in more than one sample, while several OTUs reached significant proportions of the small eukaryote community in 1 or 2 samples while they were not detected or at low proportions (< 0.05%) in other samples. Those OTUs, which are regularly detected or occasionally fairly abundant, are likely to be real members of communities and may play a non-negligible role in ecosystems. For instance, some low-frequent phylotypes may constitute a seed bank that responds to changes in environmental conditions, leading to a strong temporal dynamics that can make them detectable when particular conditions occur (Pedrós-Alió 2006; Caron & Countway 2009b; Lennon & Jones 2011). Longer surveys would give more information on low-abundance taxa ecology, testing whether most of the regularly detected OTUs happen to rise in abundance occasionally or if a large part of the rare taxa stay at low abundances.

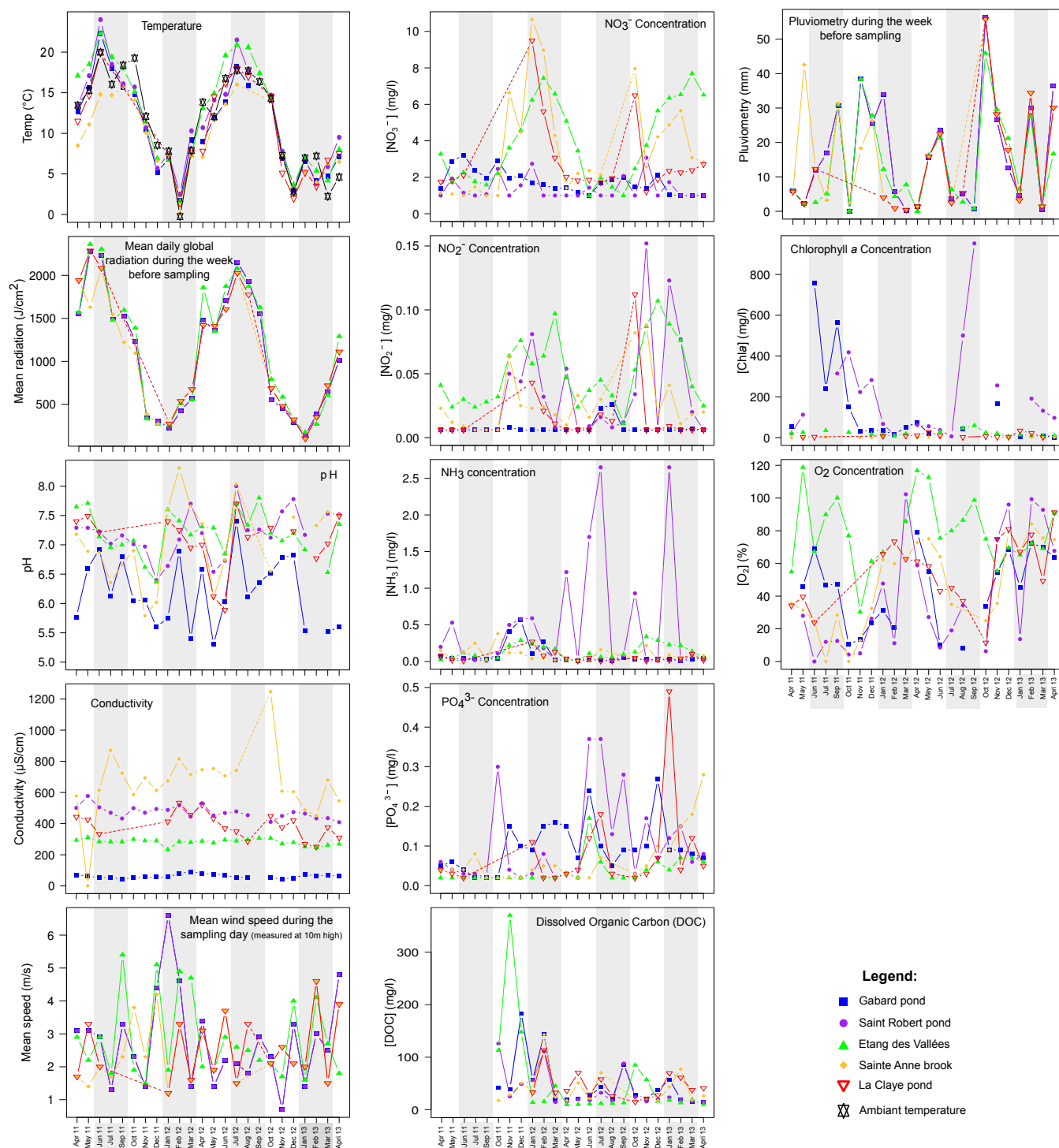
## F- Conclusions

We previously highlighted that small shallow freshwater ecosystems could be an important reservoir of microbial eukaryote diversity (Šlapeta *et al.*, 2005; Simon *et al.*, accepted). We confirm that they harbour a wide small-eukaryote diversity covering all accepted supergroups (Adl *et al.* 2012) including infrequent taxa such as glaucophytes. In addition, we show here that small-eukaryote communities are highly variable over time and space, even during a



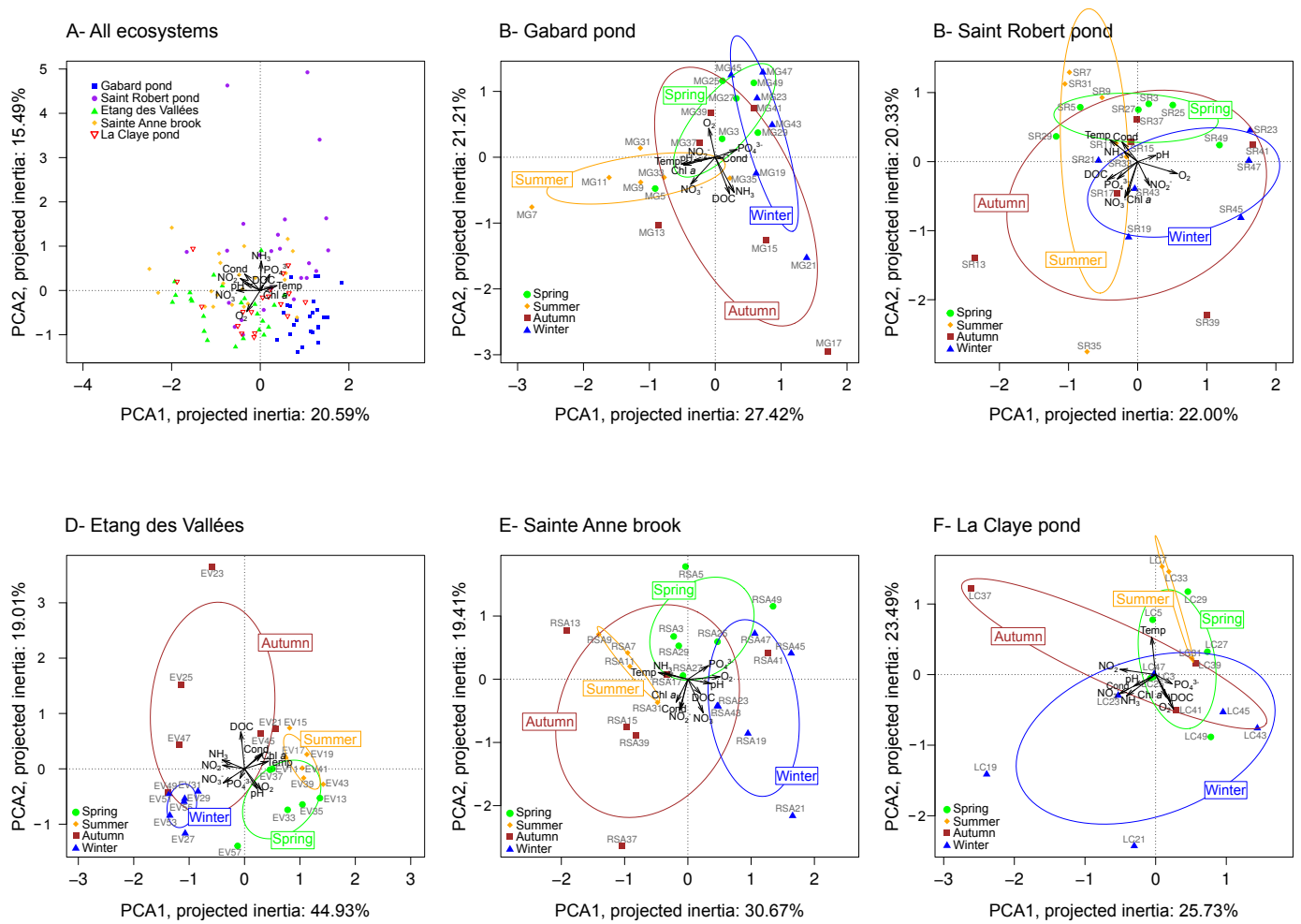
short-term survey and between geographically close systems. Since small systems are widespread, numerous and diverse, it is very likely that future temporal surveys of these understudied ecosystems will provide a better image of the eukaryotic diversity and give valuable data on infrequent phyla that could help resolving questionable relationship in the eukaryote phylogenetic tree. We were able to detect a clear seasonal dynamics of small-eukaryote communities, superimposed over more short-term variations. That seasonality was sustained by several OTUs irrespectively of their relative abundance, but we couldn't identify any seasonal trend for high-rank taxonomic groups. Longer surveys will be needed to confirm the seasonality of small-eukaryote community observed over this 2-year study.

#### **G- Supplementary information**

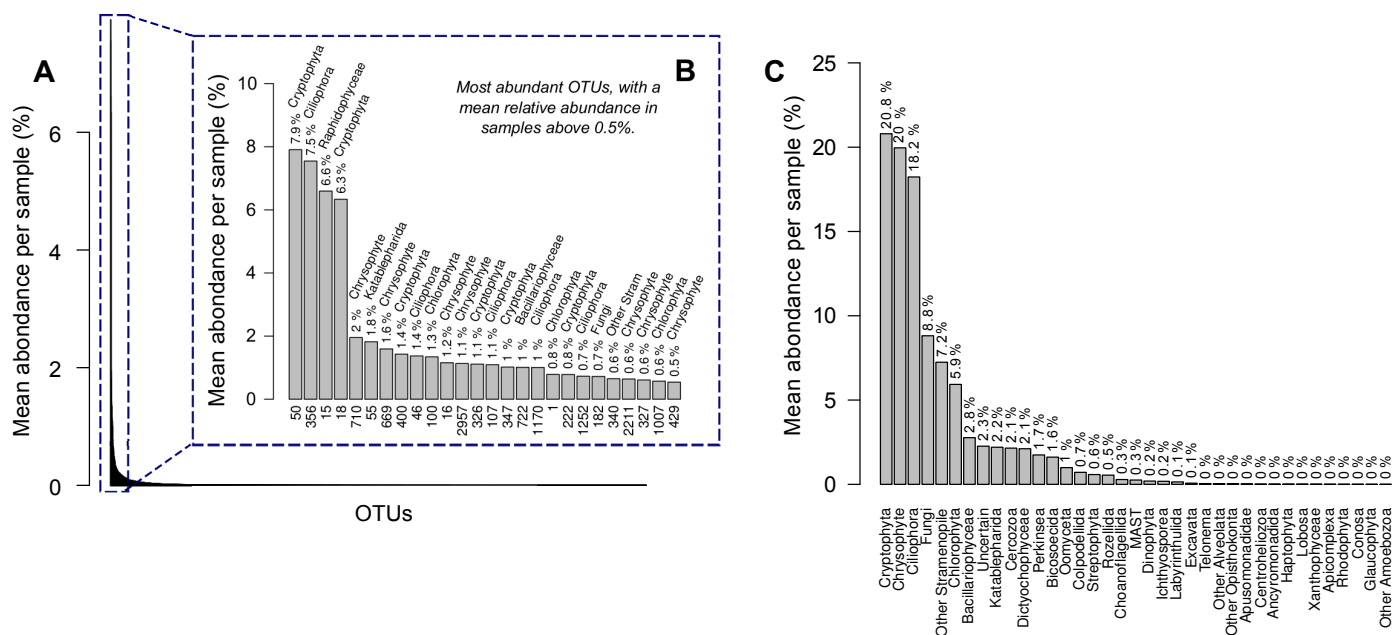


**Figure S1: Dynamics of the environmental parameters included in our analysis, over the 2-years survey.** Temperature, conductivity pH and  $\text{O}_2$  concentration were measured in situ.

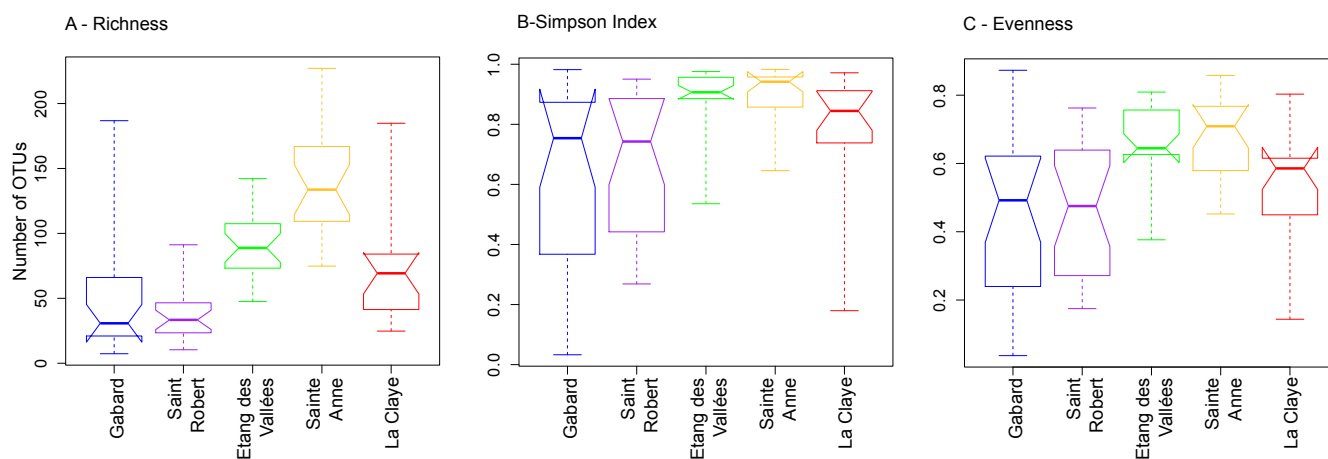
Nutrient concentrations ( $\text{NO}_3^-$ ,  $\text{NO}_2^-$ ,  $\text{NH}_3$ ,  $\text{PO}_4^{3-}$  and organic carbon) were estimated in water prefiltered through  $0.2\mu\text{m}$  pore-sized filters. Chlorophyll a concentration was measured by spectroscopy on pigments extracted from cells harvested on filters. Pluviometry, radiation, wind speed and ambient temperature (mean temperature of the week before sampling) were obtained from the closest weather stations (Météo France). Dashed lines correspond to missing data because of the drought periods: Sainte Anne brook was dried in August and September 2012, and La Claye pond from the end of July to December 2011 and in September 2012.



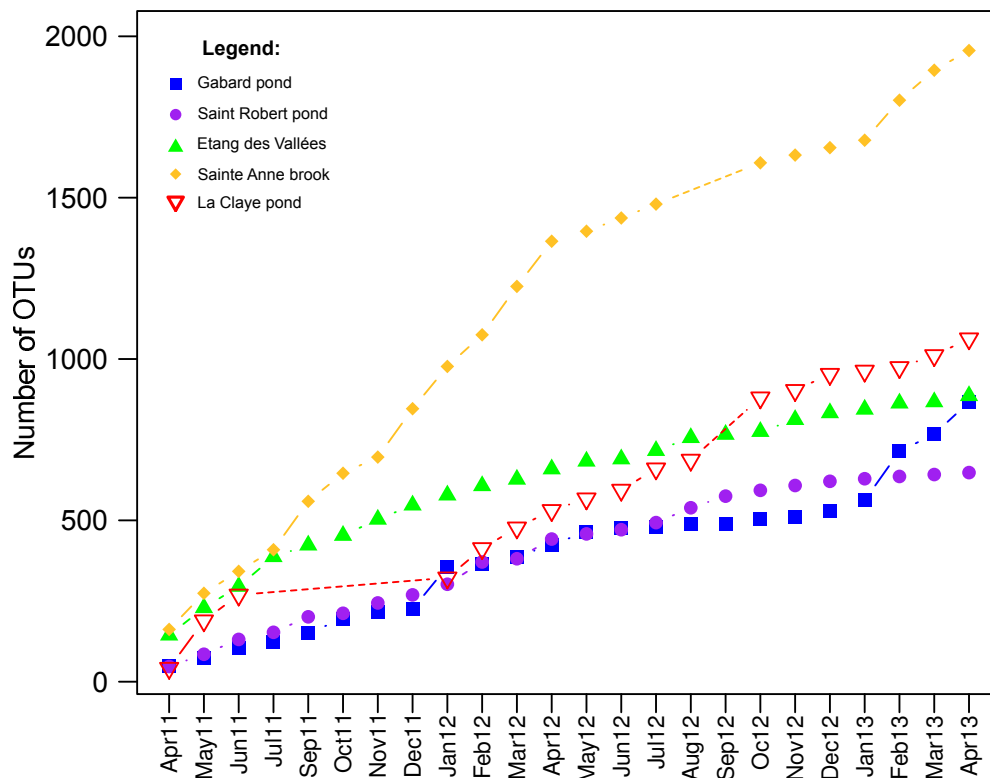
**Figure S2: Principal component analysis (PCA) plot of the physico-chemical parameters measured in the ecosystems.** PCAs were conducted for the 5 systems and 24 dates (A), or for each system separately (B to F). Samples are represented by points coloured according to the ecosystems they came from (A) or to the season when they were sampled (B to F), and their names are indicated in grey (B to F). Ellipses were drawn to show the region of the PCA where samples from the same season cluster (B to F). Cond: Conductivity, Temp: Temperature, Chl a: Chlorophyll a, DOC: Dissolved Organic Carbon.



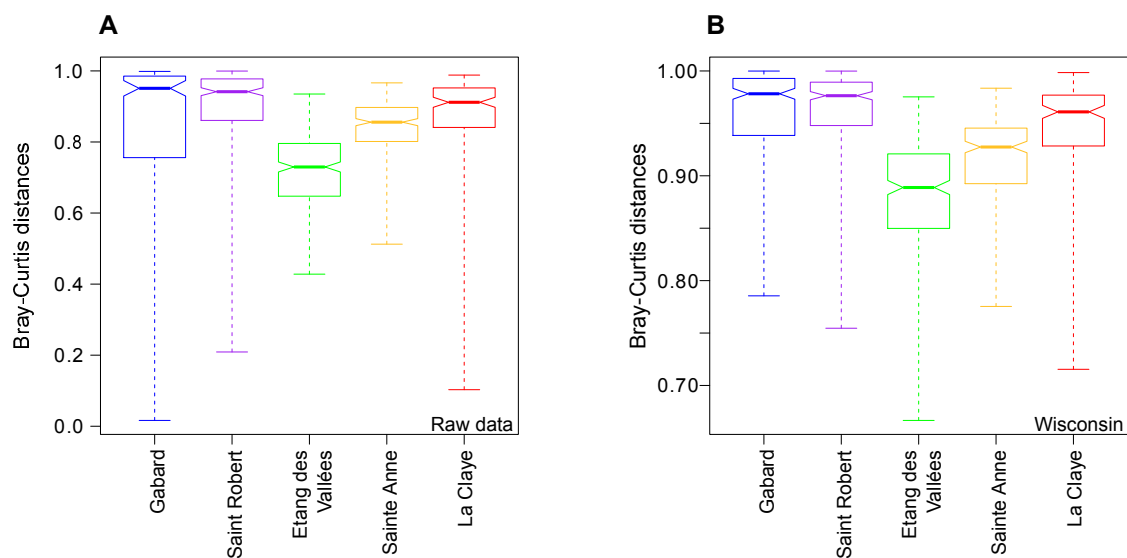
**Figure S3. General molecular diversity of the small protists in the five ecosystems over the two-year survey.** Rank-abundance curves of all OTUs (A), OTUs with a mean relative abundance in samples of at least 0.5% (B) or taxonomic groups (C), are shown. Taxonomic groups and OTUs are sorted based on their mean percentage of reads; those proportions are indicated on the top of the bars. In B, the identification number of the OTUs and their taxonomic affiliation are shown below and on the top of the bar, respectively.



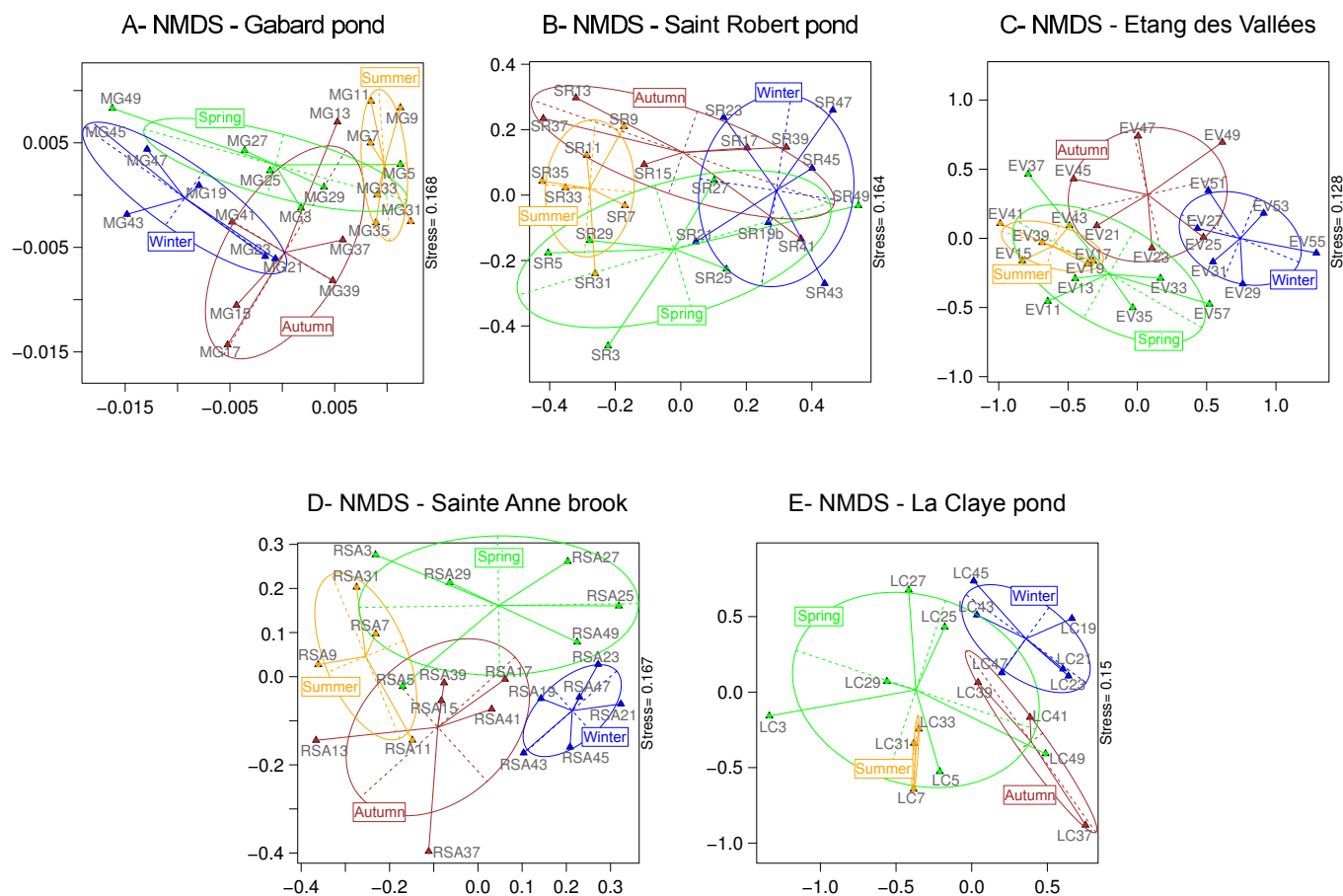
**Figure S4. Richness and diversity estimators** computed for the 24 samples from Gabard, Saint Robert ponds and Etang des Vallées, the 22 samples from Sainte Anne Brook and for the 18 samples from La Claye pond. Richness is expressed as the expected number of OTUs in random subsamples of the size of the smallest sequence library (793 reads in Gabard pond in January 2013). The thickest line inside each box represents the median on the distribution, bottom and top borders of boxes correspond to the first and third quartiles and whiskers extend to minimal and maximal distances. Notches are drawn to indicate whether medians from the five ecosystems can be considered as different (Chambers et al., 1983, p.62).



**Figure S5.** Accumulation curves showing the total number of OTUs detected in each ecosystem until the beginning of the survey. Dashed lines indicate missing data due to the drought: Sainte Anne brook was dried in August and September 2012, and La Claye pond from the end of July to December 2011 and in September 2012.



**Figure S6:** Box plot of Bray-Curtis distances between all pairs of samples from each ecosystem, calculated on raw (A) or Wisconsin standardized (B) percentage of OTUs in the samples. The thickest line inside each box represents the median on the distribution, bottom and top borders of boxes correspond to the first and third quartiles and whiskers extend to minimal and maximal distances. Non-overlapping notches are strong evidence that median of Bray-Curtis distance distributions in ecosystems differ (Chambers et al., 1983, p.62).



**Figure S7:** Non-metric MultiDimensional Scaling (NMDS) plots, based on Bray-Curtis distances computed after Wisconsin standardization and square-root transformation of all OTU frequencies in samples from Gabard (A), Saint Robert (B), Etang des Vallées (C), Sainte Anne (D) and La Claye (E). Samples are indicated in grey, and grouped by season. Stress value of each NMDS is displayed on the right border of the corresponding plot.



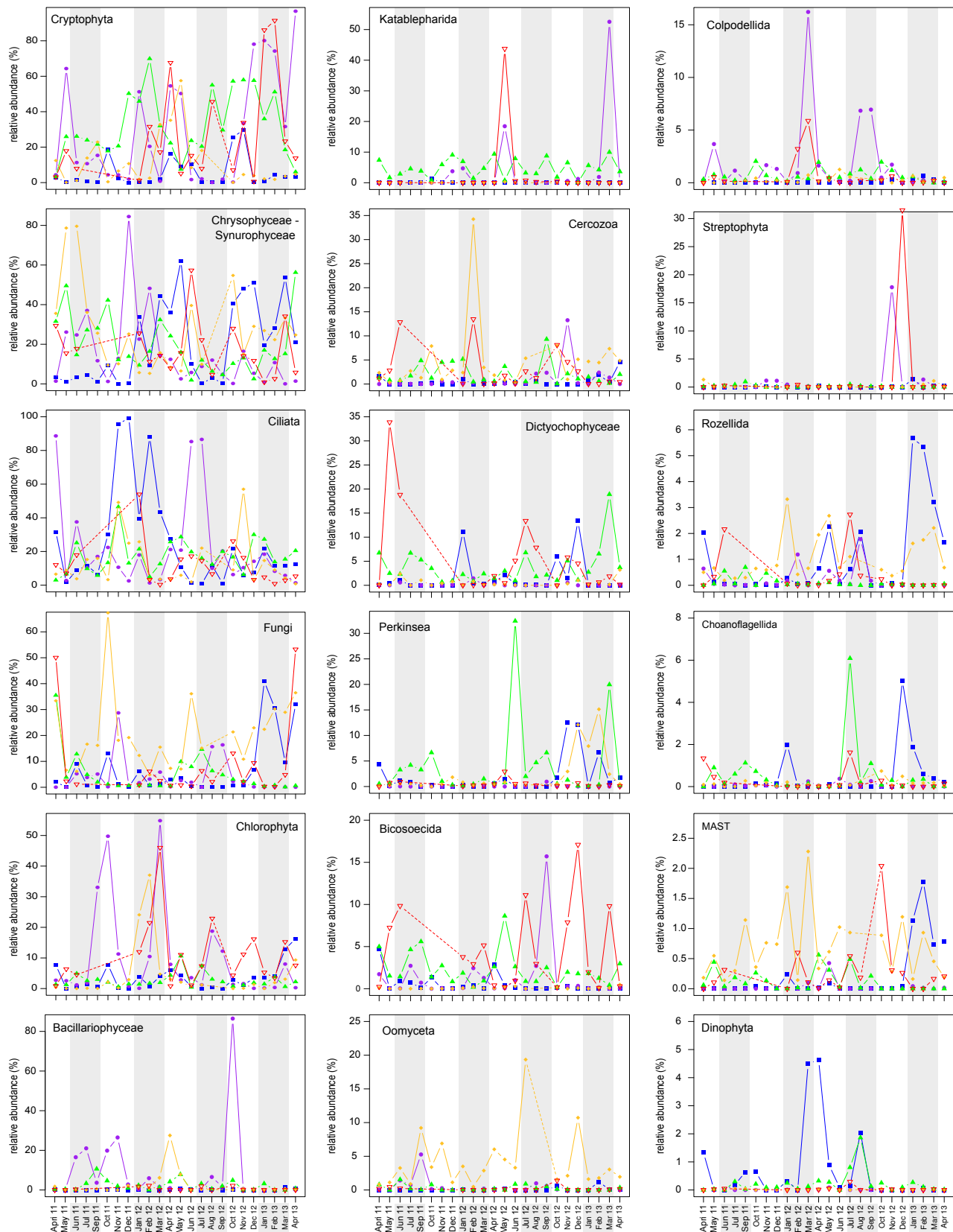
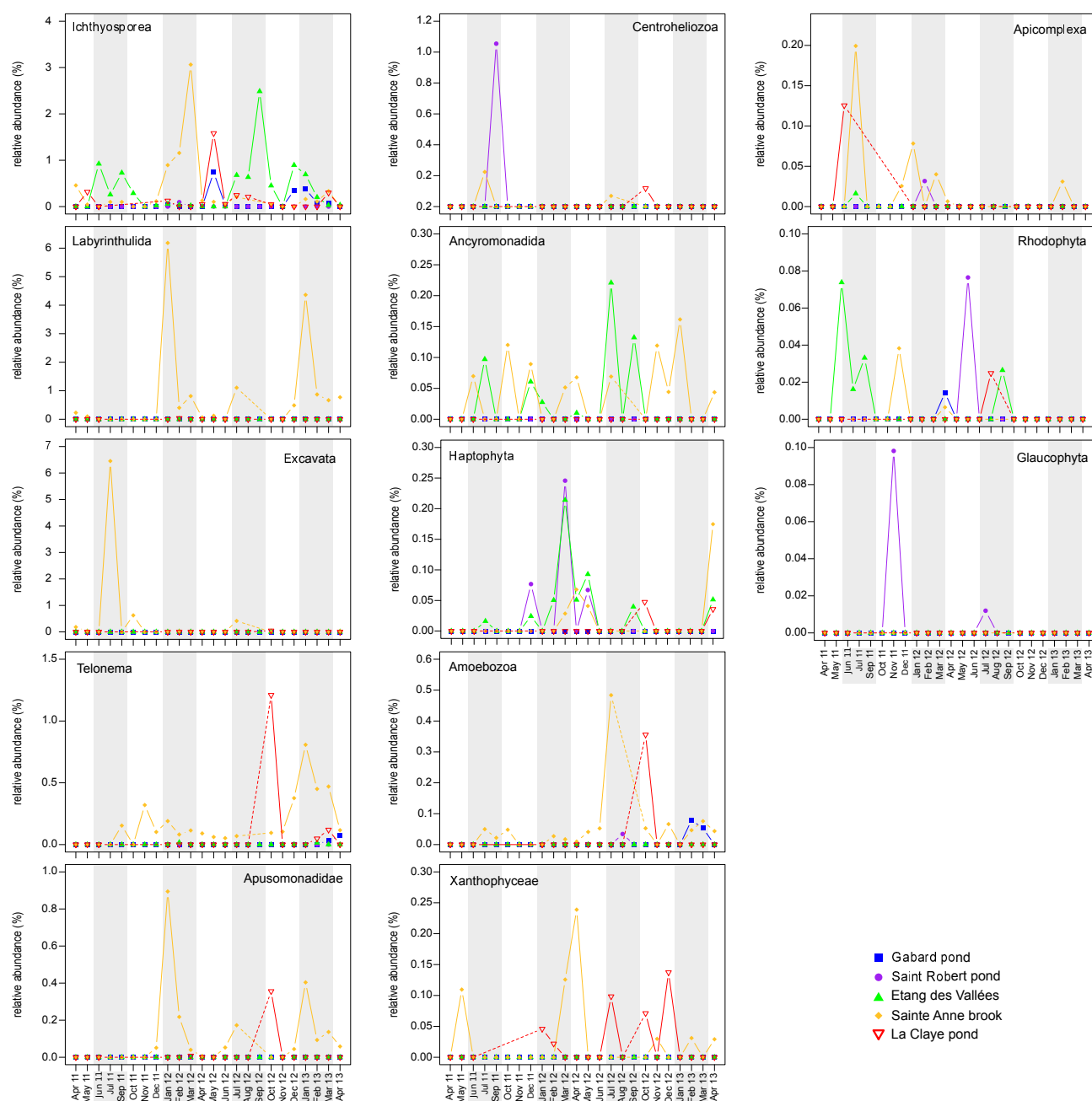


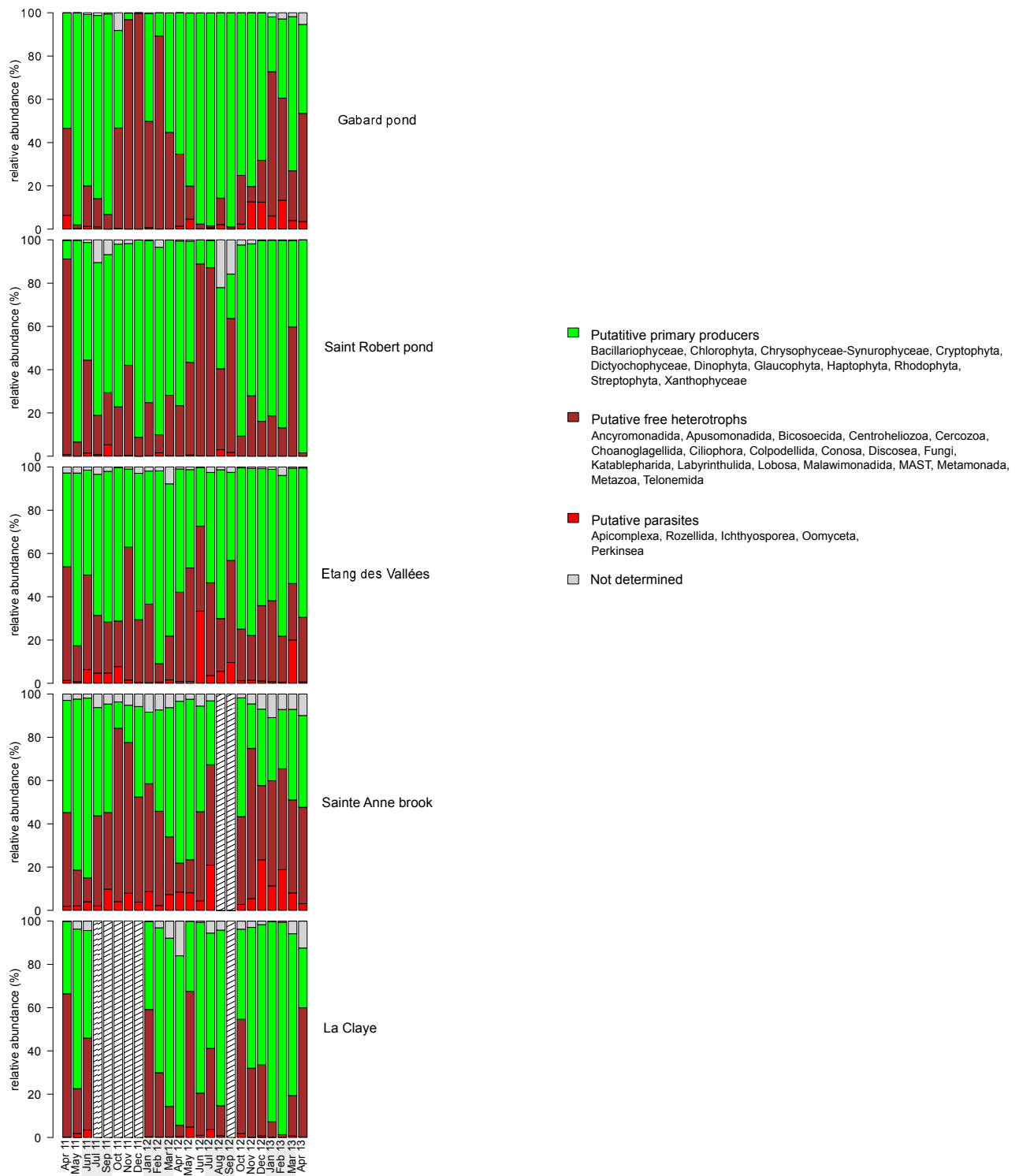
Figure S8. Part 1

- Gabard pond
- Saint Robert pond
- ▲ Etang des Vallées
- ◆ Sainte Anne brook
- ▼ La Claye pond



**Figure S8 Part 2. Temporal dynamics of detected taxonomic groups in the five sampled ecosystems.** Each color corresponds to a distinct system. Abundances are estimated as the proportion of reads affiliated to each taxa in the total number of reads in the corresponding sample. Sainte Anne brook was dried in August and September 2012, and La Claye pond from the end of July to December 2011 and in September 2012. Missing data are indicated by dashed lines.

### Chapitre 3



**Figure S9:** Histograms showing the relative abundance of putative parasites, free heterotrophs and primary producers along the two-year survey in the five sampled ecosystems. Sainte Anne brook was dried in August and September 2012, and La Claye pond from the end of July to December 2011 and in September 2012; missing data are indicated by hatched bars.

**Table S1.** Characteristics of freshwater systems studied and diversity and richness estimates for the different samples.

Sampled ecosystem	Gabard (MG)	Saint Robert (SR)	Etang des Vallées (EV)	Sainte Anne (RSA)	La Claye (LC)
Ecosystem type	Forest pond	Village pond	Large pond	Forest brook	Pond on peaty soil
GPS coordinates	48°39'15.83"N 1°55'20.26"E	48°39'54.82"N 1°56'45.28"E	48°41'23,0"N 001°54'59,2"E	48°36'45.91"N 1°58'16.61"E	48°36'31.72"N 1°56'17.33"E
Approximate surface (m <sup>2</sup> )	850 (210 x 75 m)	495 (20 x 28 m)	12 880 (210 x 75 m)	30-100 cm width at sampling point	265 (24 x 10 m)
Approximate depth (cm)	40-50	??	100-150	0-25	0-60
Water temperature <sup>2</sup> (°C)	10.8 (1.7-22.3)	11.8 (2.5-24.0)	12.6 (1.6-22.3)	9.3 (0.2-16.0)	10.0 (1.4-19.9)
pH <sup>2</sup>	6.2 (5.3-7.4)	7.2 (6.4-8.0)	7.2 (6.4-7.8)	7.0 (5.8-8.3)	7.1 (5.9-7.7)
Oxygen concentration <sup>2</sup> (%)	43.4 (8.1-78.9)	38.6 (0-102.3)	79.2 (30.4-118.8)	47.2 (0.0-84.1)	55.4 (11.5-91.2)
Conductivity <sup>2</sup> (μS/cm)	62.3 (45-91.4)	472.2 (409.0-577.0)	282.6 (234.0-312.0)	657.3 (0.6-1247.0)	389.3 (251.0-533.0)
Chlorophyll concentration <sup>2</sup> (mg/l)	132.4 (1.6-758.4)	207.8 (6.4-952.2)	21.7 (5.4-59.9)	1.3 (0.2-2.7)	10.3 (1.9-33.9)
Nitrate concentration <sup>2</sup> (mg/l NO <sub>3</sub> )	1.9 (BL-3.2)	ND	4.2 (BL-7.7)	4.1 (BL-10.7)	2.9 (1.2-9.5)
Ammonia concentration <sup>2</sup> (mg/l NH <sub>3</sub> )	0.09 (BL-0.57)	0.63 (BL-2.65)	0.14 (BL-0.34)	0.10 (BL-0.38)	0.07 (BL-0.27)
Nitrite concentration <sup>2 2</sup> (mg/l NO <sub>2</sub> )	ND	0.043 (BL-0.152)	0.051 (0.011-0.11)	0.030 (BL-0.088)	ND
DOC concentration <sup>2</sup> (mg/l)	47.2 (14.7-183.5)	37.2 (14.6-126.0)	52.3 (9.8-370.0)	40.1 (17.4-143.0)	44.9 (14.9-113.5)
Orthophosphate concentration <sup>2</sup> (mg/l PO <sub>4</sub> <sup>3-</sup> )	0.11 (BL-0.27)	0.11 (BL-0.37)	0.05 (BL-0.17)	0.07 (BL-0.28)	0.08 (BL-0.49)
Observed number of OTUs all dates taken together	867	648	886	1956	1063
Observed number of OTUs per sample <sup>2</sup>	81.2 (21-242)	70.5 (26-144)	152.1 (67-240)	257.7 (134-487)	130.0 (42-288)
Richness per sample <sup>1,2</sup>	50.3 (7.3-186.8)	38.0 (10.4-91.2)	90.2 (47.5-142.2)	138.0 (74.8-227.1)	70.6 (24.8-184.8)
Diversity: Simpson Index per sample <sup>2</sup>	0.61 (0.03-0.98)	0.67 (0.27-0.95)	0.90 (0.54-0.98)	0.90 (0.65-0.98)	0.78 (0.18-0.97)
Evenness per sample <sup>2</sup>	0.44 (0.04-0.87)	0.46 (0.17-0.76)	0.67 (0.38-0.81)	0.68 (0.45-0.86)	0.54 (0.14-0.80)

**Table S1:** Characteristics of the studied freshwater systems, and diversity and richness estimates for the different samples.

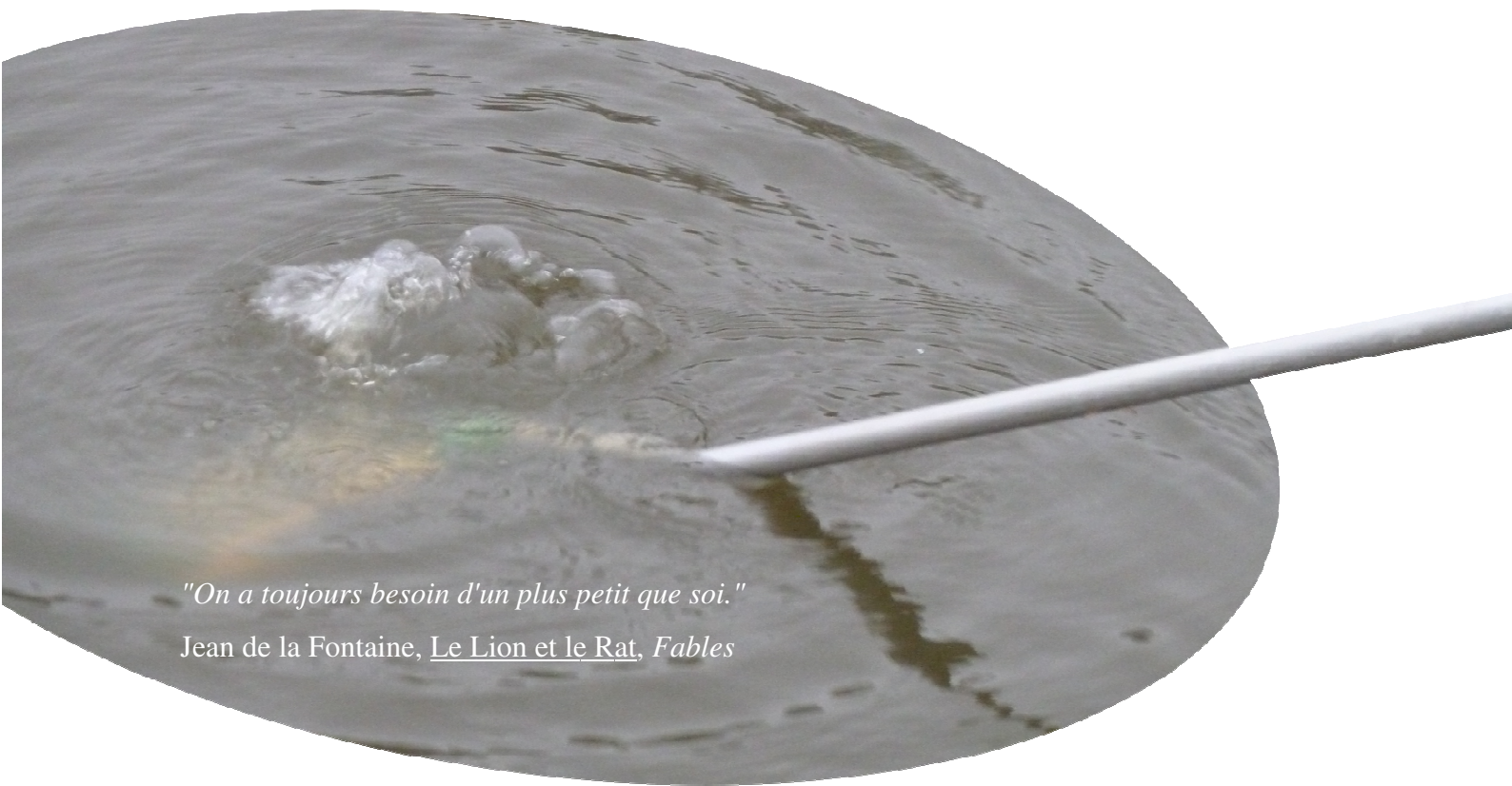
**Table S2:** OTUs detected in at least 25% of samples (i.e. 28 samples) and representing at least 0.1% of reads by sample, in mean

OTU identification number	Taxonomic affiliation	First blast hit (% identity)	Ecosystems	Mean relative abundance by sample
25	Ascomycota	Dothideomycetes_X sp. (100%)	MG, SR, EV, RSA, LC	0.30%
36	Chytridiomycota	Cladochytriales_X sp. (99.4%)	MG, SR, EV, RSA, LC	0.20%
182	Ascomycota	Leotiomycetes_X sp. (100%)	MG, SR, EV, RSA, LC	0.72%
195	Ascomycota	Dothideomycetes_X sp. (100%)	MG, EV, RSA, LC	0.14%
208	Basidiomycota	Agaricomycetes_X sp. (100%)	MG, SR, EV, RSA, LC	0.13%
1	Chlorophyta (Trebouxiophyceae)	Chlorellales_X sp. (99.2%)	MG, SR, EV, RSA, LC	0.78%
14	Chlorophyta (Trebouxiophyceae)	Choricystis sp. (100%)	MG, SR, EV, RSA, LC	0.13%
52	Chlorophyta (Chlorophyceae)	<i>Monoraphidium convolutum</i> (100%)	MG, SR, EV	0.24%
100	Chlorophyta (Chlorophyceae)	CW-Chlamydomonadales_X sp. (99.0%)	MG, SR, EV, RSA, LC	1.34%
353	Chlorophyta (Mamiellophyceae)	<i>Crustomastix</i> sp. (98.1%)	MG, SR, EV, RSA, LC	0.18%
368	Chlorophyta (Chlorophyceae)	Sphaeropleales_X sp. (100%)	MG, SR, EV, LC	0.12%
2036	Chlorophyta (Chlorophyceae)	<i>Chloromonas subdivisa</i> (99.6%)	MG, SR, RSA, LC	0.26%
158	Chrysophyceae-Synurophyceae	Chrysophyceae-Synurophyceae_XXX sp. (96.1%)	MG, SR, EV, RSA, LC	0.18%
185	Chrysophyceae-Synurophyceae	Clade-C_X sp. (97.5%)	MG, SR, EV, RSA, LC	0.49%
427	Chrysophyceae-Synurophyceae	<i>Chrysosaccus</i> sp. (98.6%)	MG, SR, EV, LC	0.23%
669	Chrysophyceae-Synurophyceae	<i>Spumella</i> _X sp. (100%)	MG, SR, EV, RSA, LC	1.59%
710	Chrysophyceae-Synurophyceae	<i>Synura petersenii</i> (99.6%)	MG, SR, EV, RSA, LC	1.95%
761	Chrysophyceae-Synurophyceae	Clade-C_X sp. (99.8%)	MG, SR, EV, RSA, LC	0.20%
2	Chrysophyceae-Synurophyceae	Clade-D_X sp. (99.2%)	MG, SR, EV, LC	0.43%
24	Dictyochophyceae	Pedinellales_X sp. (99.2%)	MG, SR, EV, RSA, LC	0.48%
91	Dictyochophyceae	Pedinellales_X sp. (99.4%)	MG, SR, EV, RSA, LC	0.15%
178	Dictyochophyceae	<i>Pseudopedinella</i> sp. (99.6%)	MG, EV, RSA, LC	0.49%
46	Ciliophora	Choreotrichia_X sp. (100%)	MG, SR, EV, RSA	1.37%
107	Ciliophora	Oligotrichia_X sp. (99.8%)	MG, SR, EV, RSA, LC	1.09%
356	Ciliophora	Stichotrichia_X sp. (99.6%)	MG, SR, EV, RSA, LC	7.55%
462	Ciliophora	Peritrichia_X sp. (100%)	MG, SR, EV, RSA, LC	0.13%
481	Ciliophora	Prostomatea_XXX sp. (99.6%)	MG, SR, EV	0.11%
7	Colpodellidae	Colpodellidae_XX sp. (99.2%)	MG, SR, EV, RSA, LC	0.18%
20	Cercozoa	Novel-clade-10_XXX sp. (99.8%)	MG, SR, EV, RSA, LC	0.31%
18	Cryptophyta	Cryptomonadales_X sp. (99.4%)	MG, SR, EV, RSA, LC	6.34%
50	Cryptophyta	Cryptomonadales_X sp. (100%)	MG, SR, EV, RSA, LC	7.91%
65	Cryptophyta	Basal_environmental_Cryptophyceae_X sp. (99.8%)	SR, EV, RSA, LC	0.50%
142	Cryptophyta	Cryptomonadales_X sp. (99.8%)	SR, EV, LC	0.39%
326	Cryptophyta	<i>Komma caudata</i> (99.4%)	SR, RSA, LC	1.11%
347	Cryptophyta	<i>Cryptomonas obovoidea</i> (100%)	MG, SR, EV, RSA, LC	1.02%
400	Cryptophyta	Cryptomonadales_X sp. (100%)	SR, EV, LC	1.43%
55	Katablepharidophyta	Katablepharidales_XX sp. (99.8%)	MG, SR, EV, RSA, LC	1.82%

**Table S2:** OTUs detected in at least 25% of samples (i.e. 28 samples) and representing at least 0.1% of reads by sample, in mean.

## Chapitre 4 : Résilience rapide des communautés planctoniques de petits eucaryotes après sécheresse complète de leur écosystème.

---



*"On a toujours besoin d'un plus petit que soi."*

Jean de la Fontaine, Le Lion et le Rat, *Fables*





## I. Introduction : Contexte et objectifs

Nous avons initialement choisi les cinq écosystèmes dans lesquels nous avons suivi les communautés planctoniques de petits eucaryotes pendant deux ans parce qu'ils étaient censés être permanents. Il s'est par la suite avéré que ce n'était pas tout à fait le cas... Le volume d'eau contenu dans les mares, surtout dans La Claye, et le débit du ruisseau étaient extrêmement variables. Notamment, la première année d'échantillonnage a connu un été particulièrement sec, qui s'est traduit par l'assèchement de la mare de La Claye pendant une période couvrant 5 mois d'échantillonnage. Au cours de la seconde année, cette mare n'a été asséchée qu'un mois, en septembre 2012, tandis que le ru Sainte Anne était complètement sec (même les flaques disjointes s'étaient asséchées) en août et septembre de cette même année.

L'assèchement total d'un écosystème aquatique, même de durée limitée, constitue certainement un événement particulièrement stressant pour les organismes planctoniques qui y vivent. En effet, ils perdent leur habitat, subissent des chocs hydriques et les chocs osmotiques associés, à la fois lors de l'assèchement et de la remise en eau. On peut donc supposer que de tels événements aient un impact important sur les communautés de petits eucaryotes planctoniques dans leur ensemble, à court ou plus long terme, notamment si une toute nouvelle communauté se forme à partir d'organismes immigrants après la remise en eau.

Nous avons profité de l'opportunité que nous proposaient ces sécheresses inattendues pour tenter d'estimer cet impact. Dans le présent chapitre, nos principaux objectifs visaient à :

- Comparer la diversité et la variabilité spatio-temporelle des communautés planctoniques avec celles des sédiments.
- Estimer la capacité de résilience des communautés planctoniques après un événement de sécheresse.

L'échantillonnage n'avait pas initialement été prévu pour répondre à de telles questions, et n'est donc pas particulièrement adapté. Il aurait fallu pouvoir comparer communautés des sédiments et de l'eau de surface aux mêmes dates, avec un nombre équivalent et suffisant d'échantillons dans chaque cas, sans préfiltration des échantillons d'eau. Certaines hypothèses auraient sans doute été plus solides avec un échantillonnage plus fréquent autour des périodes d'assèchement et de remise en eau. Cependant, les premières observations, présentées dans l'ébauche d'article ci-après, permettent de proposer des hypothèses qui pourront être testées par la suite, et indiquent déjà clairement la forte et rapide résilience des communautés planctoniques d'eucaryotes après un assèchement total de leur écosystème.

**II.       Manuscrit de l'article 4 : Fast recovery of planktonic eukaryote communities after drought events in two shallow freshwater ecosystems.**

**Fast recovery of planktonic eukaryote communities after drought events in two shallow freshwater ecosystems**

Marianne Simon, Ludwig Jardillier, Philippe Deschamps, David Moreira, Gwendal Restoux,  
Paola Bertolino, Purificación López-García

*Manuscript in preparation*

## A- Abstract

Sudden and drastic changes in environmental conditions are important challenges to microbial communities in little buffered ecosystems. The capacities to resist change or to recover after environmental perturbations ensure the stability of the ecosystem functioning on the long run. Small and shallow freshwater ecosystems undergo rapid and drastic shifts in their environmental conditions and may experience seasonal droughts. We were able to monitor such a drought event during a temporal survey of protist diversity carried out monthly for two years and based on massive sequencing of 18S rRNA gene amplicons in one brook and one pond in temperate latitudes. The eukaryote assemblages in the exposed bottom sediment were rich and highly diverse, but the community composition was very different than those of planktonic communities prior to the drought event. When the water level was recovered, the two ecosystems recovered planktonic protist community structures and composition very similar to those occurring prior to the drought event within a month. This rapid recovery indicates that these communities present a high resilience to drought events, a property that could be essential in small and shallow freshwater ecosystems, which frequently undergo this kind of perturbation. This also suggests a predominance of taxa able to produce resting stages (e.g. cysts) that germinate or recover growth rapidly after favourable conditions are restored.

### Keywords:

18S rDNA / freshwater / microbial eukaryotes / plankton / drought / resilience

## B- Introduction

Microorganisms, both prokaryotes and eukaryotes, are key players in aquatic ecosystem functioning (Cho & Azam 1988; Massana 2011). In particular, protists are largely involved in biogeochemical cycles through nutrient cycling (Caron 1994), CO<sub>2</sub> fixation (Jardillier *et al.*, 2010) and predation (Zubkov & Tarran 2008), and are characterised by a great molecular diversity (e.g. Richards & Bass 2005). Although far from being completely apprehended, the diversity of the microbial compartment in marine and, to a lesser extent, in freshwater ecosystems has been studied in past years by a variety of molecular tools, showing highly complex communities (e.g. Richards & Bass 2005; López-García & Moreira 2008; Massana & Pedrós-Alió 2008; Triadó-Margarit & Casamayor 2012; Taib *et al.* 2013). Much less is understood about how marine and freshwater microbial communities change through time, although it has been observed that microbial eukaryote communities succeed each other over seasons leading to permanent reassembling (e.g. Bamforth 1958; Kim *et al.* 2014). These changes are likely due to bottom-up and top-down biotic factors as well as physico-chemical environmental parameters (e.g. Lepère *et al.* 2006; Lara *et al.* 2011). Seasonal changes occur regularly and microbial species are adapted to (and often depend on) those changes. However, little is known about microbial eukaryote communities' answers to strong perturbation events, such as droughts which must be one of the most drastic stresses that may occur for aquatic microbes since it can lead to the desiccation of their natural habitat.

Small and shallow freshwater ecosystems such as shallow lakes, ponds or streams are little buffered, and thus particularly sensitive to drought periods that may cause dryness of the whole system. Indeed, these aquatic systems are highly dynamic (Angélibert *et al.* 2004), with a variable water content and a relatively short lifespan (Downing 2010), since they are small depressions in soil that are usually mainly filled by rains. The functioning and ecology of these particular systems remains poorly investigated and understood although they are widespread, numerically very important and contribute significantly to global biogeochemical cycles (Downing 2010). Moreover, a few studies revealed that they harbour a wide microbial eukaryote diversity (Šlapeta *et al.* 2005; Simon *et al.*, accepted). Because of the global warming observed since the late 19<sup>th</sup> century, strong meteorological episodes such as droughts are likely to rise in frequency and strength in the near future, (IPCC 2007). Considering the global importance of protist activity and their high diversity and abundance in those widespread ponds and brooks, it is crucial to understand how droughts affect microbial eukaryote communities of the small and shallow freshwater ecosystems.

During a temporal survey of planktonic protist diversity that we carried out in a previous work (Simon et al, in prep, see Chapter 3), two of the systems under study, a brook and a pond, experienced long (ca. 5 months) and short (ca. 1 and 2 months) desiccation events. This offered us the possibility to see how the communities recovered and whether they exhibited high resilience or whether very different communities established after the systems were refilled with water. In this work, we compared the protist molecular diversity determined by 454-pyrosequencing of amplified 18S rRNA gene fragments in water, and in sediment when the systems were dried.

## C- Materials and Methods

### *Study sites and sampling*

Samples had been collected monthly, from April 2011 to April 2013, in La Claye pond and Sainte Anne brook. These two semi-permanent small and shallow freshwater systems are located in the Natural Regional Park of the Chevreuse Valley (France, South of Paris). These sites and the collection of surface samples have been described in Simon et al. (in press). Briefly, water was collected at surface using sterile plastic bottles and processed immediately back to the laboratory. Planktonic cells were collected onto 0.2 µm pore-size Nucleopore filters (Millipore) after a prefiltration step through 5 µm pore-size Nucleopore membranes (Whatman). Filters were then stored frozen until DNA extraction. During drought periods, samples were collected at the surface of the sediment using 50 ml sterile plastic tubes (Becton Dickinson Biosciences), then transferred to 5 ml cryotubes back to the laboratory and immediately frozen until DNA extraction.

### *DNA extraction, amplification and sequencing of 18S rDNA fragments*

DNA was extracted from 0.25 g of sediment using the PowerSoil DNA extraction kit (MoBio) according to the manufacturer's instructions. DNA was eluted in 80 µl of elution buffer (Tris 10 mM, pH 8). 18S rDNA fragments including the V4 hypervariable region were amplified using primers EK-565F (5'-GCAGTTAAAAAGCTCGTAGT) (Simon *et al.* accepted) and 18S-EUK-1134-R-UNonMet (5'-TTTAAGTTTCAGCCTTGCG) biased against Metazoa (Bower *et al.* 2004). Primers were tagged with different Molecular IDentifiers (MIDs) to allow multiplexing and later differentiation of PCR products from the nine sediment samples with amplicons from planktonic samples presented in a previous study (Simon *et al.* in prep, see Chapter 3). PCR amplifications were conducted in a total reaction volume of 25 µl, using



1.5 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP (PCR Nucleotide Mix, Promega), 0.3 μM of each primer, 1 to 2.5 μl of DNA and 0.5 U HotStart Taq polymerase (Taq Platinum, Invitrogen). Amplification was realized during 25 cycles (94°C for 30 s, 58°C for 45 s and 72°C for 90 s), after 3 min of denaturation at 94°C and before a final extension step at 72°C for 10 min. Amplicons from 5-8 independent PCR amplifications for each sample were pooled together. Each pool was then purified using the QIAquick PCR purification kit (QIAGEN), according to the manufacturer's instructions. Finally, the same amount of purified PCR products from sediment samples were pooled along with amplicons from planktonic samples and pyrosequenced using the 454 GS-FLX Titanium technology from Roche (Beckman Coulter Genomics). Two distinct 0.25 g sediment samples from La Claye pond collected in December 2011 were treated independently from DNA extraction to sequence analysis to assess the reproducibility of our method for sediment samples. We thus processed a total of 9 samples: 5 from La Claye pond collected from the end of July to December 2011 plus one replicate from the latter sampling month, one from La Claye in September 2012 and two from Sainte Anne brook in August and September 2012.

#### *454 pyrosequence analysis*

A total of 48,429 pyrosequences were obtained from the nine sediment samples. We applied a series of filters to keep only high-quality reads. To begin with, sequences with errors in the primer region and / or positions with undetermined bases were eliminated using a local pipeline (Bachy *et al.* 2013; Simon *et al.* accepted). Quality-checked pyrosequences were analyzed with AmpliconNoise (Quince *et al.* 2011) integrated to our local pipeline in order to eliminate PCR and 454 sequencing errors. After filtering, 34,821 correct reads obtained from sediment samples were retained. They were considered together with correct reads obtained from planktonic samples of the same systems but between drought periods (Simon *et al.* in prep, see Chapter 3), to build Operational Taxonomic Units (OTUs) using AmpliconNoise. OTUs were composed of clustered filtered reads, with a 98% identity similarity cut-off. For more precaution, OTUs containing only one sequence were eliminated. The most abundant read of each OTU was used as a reference and blasted against the PR2 database (Guillou *et al.* 2013) for taxonomic assignation of the OTUs. Sequences in all OTUs were then attributed to their sample according to their MIDs. Chimerical OTUs were eliminated during a stringent procedure combining both manual and automated steps (Simon *et al.*, in press). OTUs affiliated to cyptophyte nucleomorphs were not included in the analysis.

### Statistical analysis

Statistical analysis were all conducted using the R software (R Development Core Team 2013). Diversity and richness estimates were computed based on raw counts of reads attributed to each OTU, using the R package 'Vegan'. Richness was estimated by rarefaction, as the expected number of OTUs in a random subsample of each sequence library, having the size of the smallest library (Hurlbert 1971). Diversity was estimated by the Simpson index, calculated as  $D' = 1 / \sum_{i=1}^S (f_i^2)$  (Simpson 1949) and evenness as  $e = - \sum_{i=1}^S f_i \ln(f_i) / \ln(S)$

(Pielou 1966) with S being the observed number of OTUs and  $f_i$  the frequency of each OTU<sub>i</sub> in the sample. To evaluate overall differences between eukaryotic assemblages, pairwise Bray-Curtis dissimilarities were calculated between all samples  $(\sum_{i=1}^S |x_{ji} - x_{ki}| / \sum_{i=1}^S (x_{ji} + x_{ki}))$ , with  $x_{ji}$  and  $x_{ki}$  the abundances of OTU<sub>i</sub> in samples j and k, respectively, and S the number of OTUs observed in libraries j and k), based on OTU percentages of reads (instead of raw counts to not consider differences due to different numbers of reads). They were computed using the R 'Vegan' package (Oksanen *et al.* 2013). The same package was used to draw Non metric MultiDimensional Scaling (NMDS) plots comparing communities from the two-year survey of both ecosystems or on La Claye pond and Sainte Anne brook separately. They were based on Bray-Curtis dissimilarities calculated after square-root transformation and Wisconsin standardization (Bray & Curtis 1957) of OTU percentages. Ellipses were drawn on NMDS plots using the R package 'Ade4' (Dray & Dufour 2007) to highlight whether the communities were collected in water or sediment. Boxplots were drawn with notches to indicate whether the medians of the represented distributions could be considered as different (Chambers *et al.* 1983).

## D- Results and discussion

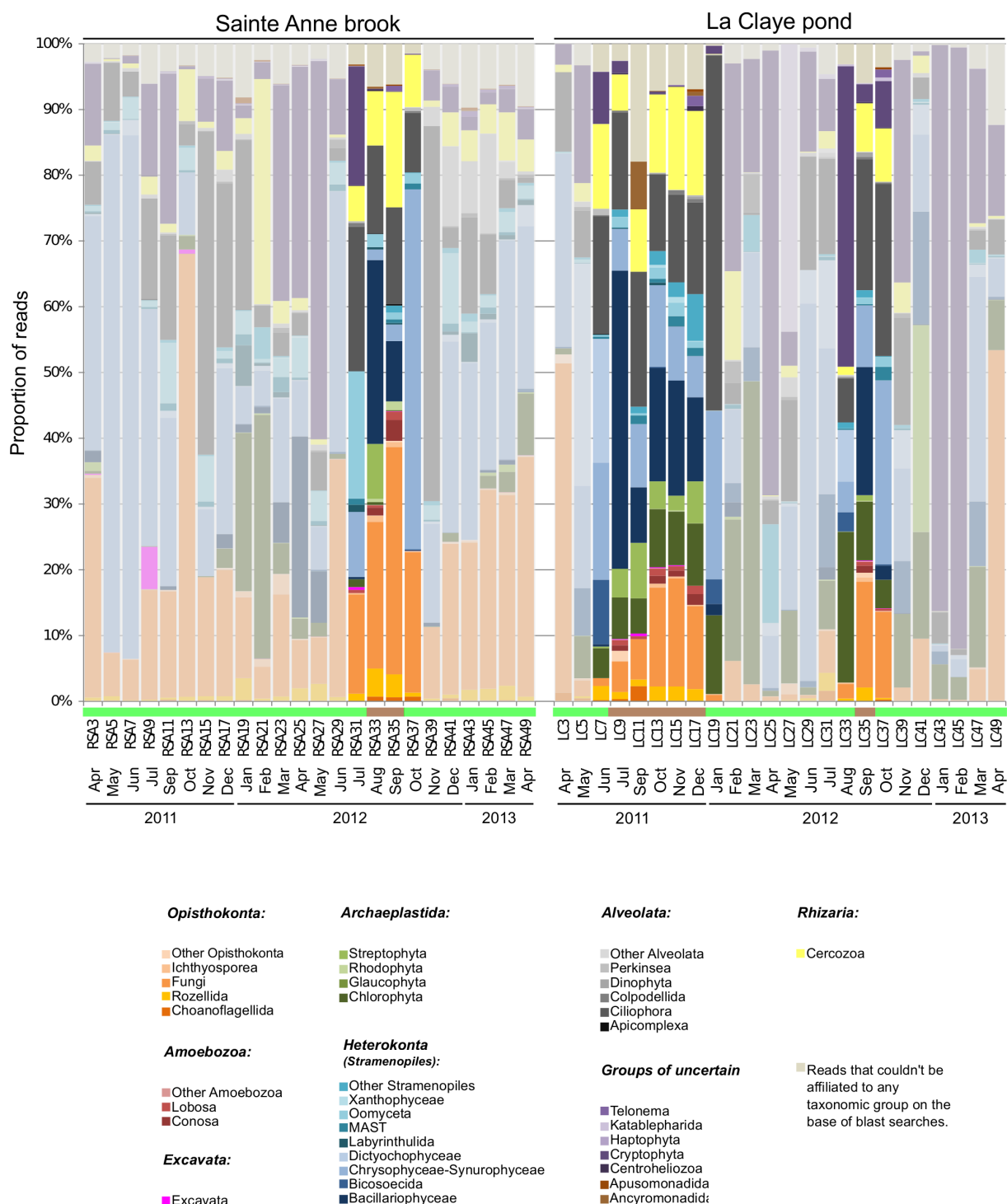
In a previous study, the diversity of microbial eukaryotes was studied over a two-year survey with a monthly sampling in a small brook and a shallow pond (Simon *et al.*, in prep, see Chapter 3). Two of the studied systems underwent desiccation during drought periods. La Claye pond was found dried out during five sampling months, from the end of July 2011 to the beginning of December 2011 and then again in September 2012. The Sainte Anne brook dried out only during the second sampling year, in August and September 2012.

To monitor how communities changed during the periods of desiccations, we analysed protist diversity from surface sediment collected in the dried bed of the pond and the brook (called sediment communities herein) based on 454-pyrosequenced rDNA fragments, and then compare that diversity to planktonic communities previously described (Simon et al., in prep, see Chapter 3) from the same systems but when they were flooded.

We made two replicates for the sediment sample collected in La Claye pond in December 2011, which were analysed independently. The community compositions detected in both replicates were very similar, as revealed by the Bray-Curtis dissimilarity of 0.33 as compared to the mean pairwise dissimilarity between all pairs of sediment communities from La Claye pond of 0.68. During the next analyses, only one replicate was included.

#### *Global composition of sediment communities from the bed of dried systems*

Over the two-year survey, we detected 3,132 OTUs out of which 1,176 were recorded in the 8 sediment samples and 2,560 in the 40 water samples. After taxonomic affiliation of OTUs based on sequence similarities, all recognized supergroups (i.e. Stramenopile-Alveolata-Rhizaria (SAR), Archaeplastida, Excavata, Amoebozoa, Opisthokonta (Adl *et al.* 2012) and several groups of uncertain affiliation were detected both in sediment and water habitats, occasionally in very different proportions (Figure 1).

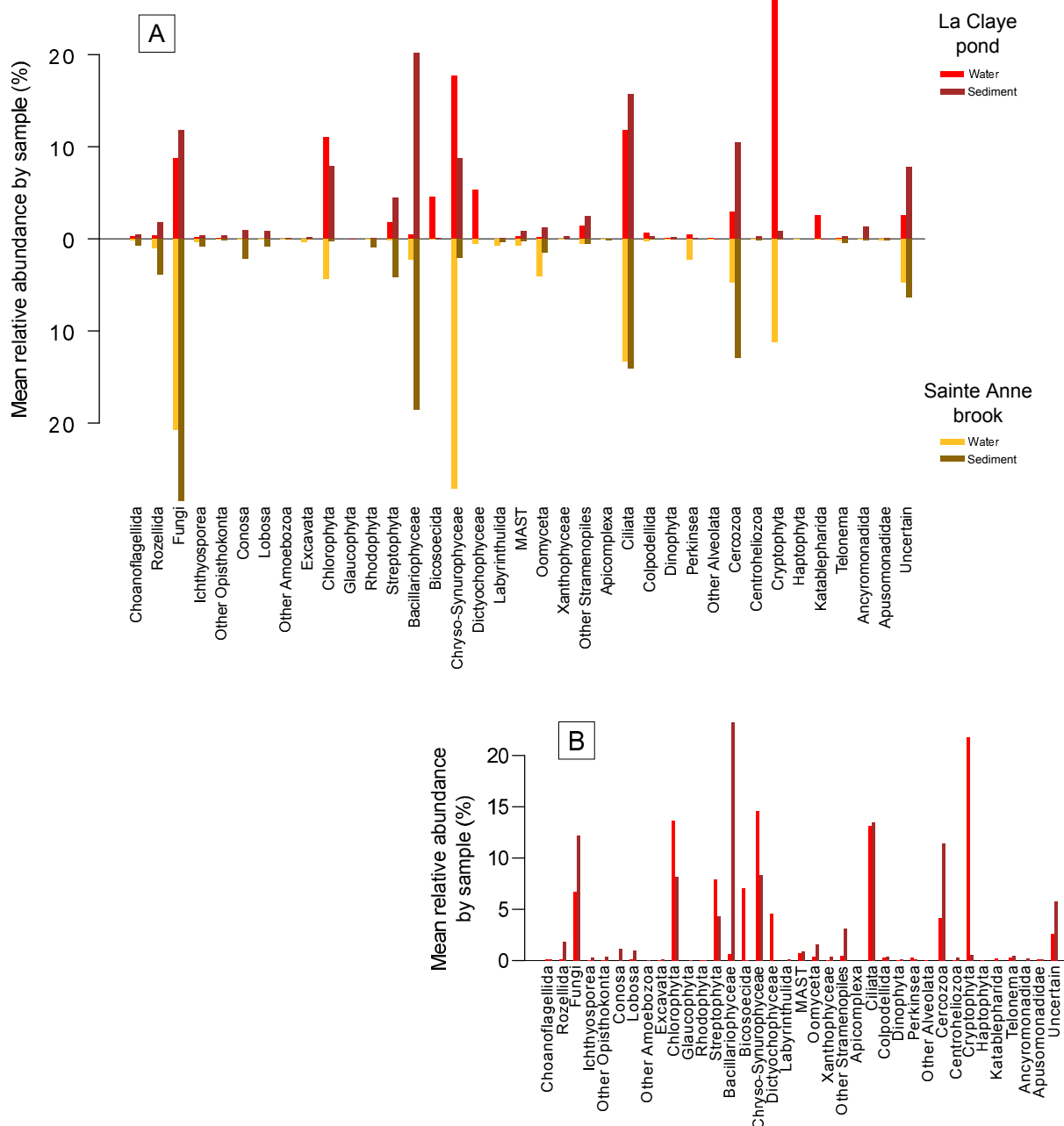


**Figure 1: Taxonomic composition of eukaryotes of the two-year survey in Sainte Anne brook and La Claye pond.** The horizontal bar below the histograms indicate whether the vertical bar above correspond to a water (green) or sediment (brown) community.

Stramenopiles constituted the most abundant supergroup in sediment samples from La Claye pond (33.7% of the reads per sample on average) (Figure 1 and 2A), and the second

most abundant in the two samples from the brook (23.3% of the reads per sample on average). That large abundance is in accordance with previous records of this supergroup in a wide diversity of environments, from marine and freshwater systems (Richards *et al.* 2005; Lefèvre *et al.* 2008; Triadó-Margarit & Casamayor, 2012; Kirkham *et al.* 2013; Simon *et al.*, accepted) to soils (e.g. Moon-van der Staay *et al.*, 2006). Chrysophyceae-Synurophyceae and above all Bacillariophyceae were the most abundant stramenopiles of sediment assemblages, but OTUs affiliated to Labyrinthulida, MAST groups, Oomyceta and Xanthophyceae were also detected. In Sainte Anne brook, Opisthokonta constituted the major supergroup, with 33.9% of reads per sample (average from the two samples) affiliated to that supergroupe. They were mostly composed of fungi (28.4% of reads per samples on average), which are an important component of aquatic, sediment and soil habitats (Moon-van der Staay *et al.* 2006; Lefèvre *et al.* 2008; Taib *et al.* 2013). However, OTUs affiliated to Rozellida-Cryptomycota were quite abundant too (3.9% of reads per sample on average), and Ichthyosporea and Choanoflagellata were also detected. Opisthokonts were less abundant in sediment from La Claye pond (14.7% of reads per sample on average), but still constituted the third most abundant group in that system. Interestingly, the greatest abundance of opisthokonts in the brook than in the pond was also detectable in planktonic communities (Figure 1 and 2A). Alveolates are also common members of freshwater lakes and ponds (Šlapeta *et al.* 2005; Taib *et al.* 2013; Simon *et al.*, accepted) and were always highly abundant in sediment samples from both the brook and the pond (Figure 1). In both cases, they were mainly represented by ciliates (Figure 1 and 2A), which are known for their high abundance and widespread distribution in soils (Foissner 1997), freshwaters (e.g. Sonntag *et al.* 2006; Lefèvre *et al.* 2008; Charvet *et al.* 2012; Mangot *et al.* 2012) and sediments (López-García *et al.* 2003). OTUs affiliated to Archaeplastida (mostly Streptophyta and Chlorophyta) were also abundant in all sediment samples from La Claye and in August 2012 in Sainte Anne brook (Figure 1). Rhizaria, exclusively represented by OTUs affiliated to Cercozoa, were abundant in all sediment sample with no exception (Figure 1), representing 10.4 and 12.9% of reads per sample from La Claye pond and Sainte Anne brook respectively, on average. Amoebozoans represented as much as 1.8 and 3.0% of reads per sample from La Claye pond and Sainte Anne brook respectively, on average. Apusozoans were generally low abundant, but reached 7.3% of reads from the sediment community sampled in La Claye pond in September 2011. In addition, OTUs affiliated to excavate, cryptophytes, katablepharids (only in La Claye pond), telonemids and centrohelids were also detected in sediment samples, though in lower abundance. So, eukaryote assemblages from sediment of the dried up bed of studied systems

are highly diverse. They may be composed of sediment resident organisms and usually planktonic cells that sedimented during the drying up.

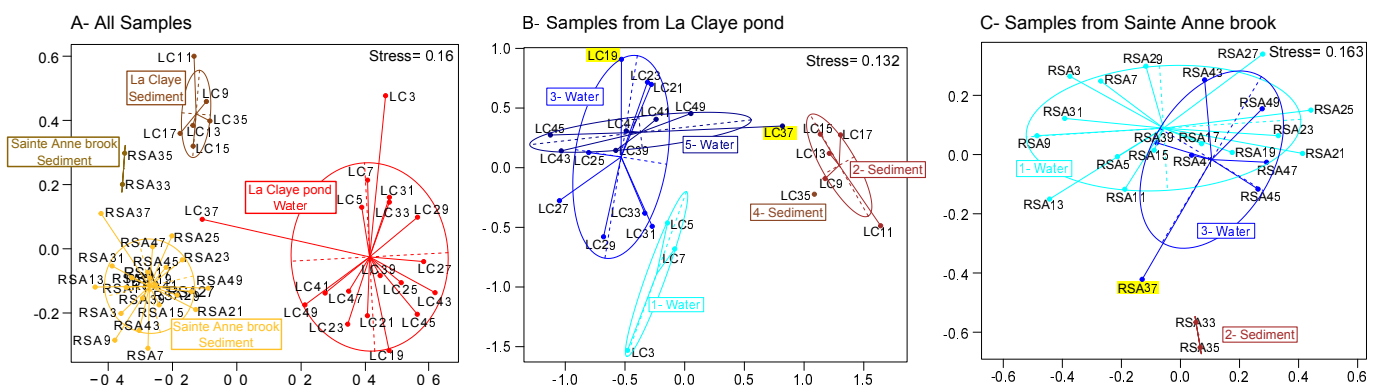


**Figure 2: Global taxonomic composition of sediment and water communities in Sainte Anne brook and La Claye pond.** Mean relative abundances per sample were calculated on all samples (A) or on four sediment samples from summer and autumn 2011 and four water samples from the same period in 2012 from La Claye pond (B).



*Distinct communities in sediments from the bed of dried systems and surface water*

Communities inhabiting the dry sediment during the drought periods were always different from planktonic communities living in the water mass (Figure 3). Sediment and planktonic communities were less different in Sainte Anne brook than in La Claye pond, possibly due to mixing of the sediment and water. The brook is indeed very shallow with a water column <20 cm in height, leading sediments to be easily resuspended either by the running water itself, by the activity of wild boars which use this brook as watering area (numerous traces are found) or by the sampling process despite all the precautions we took to avoid this possible bias.



**Figure 3:** *Non Metric MultiDimensional Scaling (NMDS) plots grouping eukaryote communities of the two-year survey in Sainte Anne brook and La Claye pond (A), La Claye pond only (B) and Sainte Anne brook only (C). In B and C, ellipses group eukaryote communities of the same periods defined by droughts. The first water communities following drought periods are highlighted in yellow.*

First, planktonic and sediment eukaryote assemblages differed by their taxonomic composition. Bacillariophytes were clearly the dominant stramenopile group in the sediments (19.7% of the reads per sample, on average) but not in the water habitat where they were more than 13 times less abundant (1.5% of the reads per sample, on average) (Figures 2A), which may reflect their ability to form resting stages and include many diatoms with a benthic life (e.g. McQuoid *et al.* 2002; Kamp *et al.* 2011). Members of Rozellida-Cryptomycota have been reported from a large range of habitats, namely freshwater, soil, peat, and marine sediment (Lara *et al.* 2010; Gleason *et al.* 2012). They were especially abundant in the sediment communities in our two systems (3.9 and 1.7% of the reads per sediment sample in Sainte Anne and La Claye, respectively, on average), rather than in planktonic assemblages (1 and 0.1% of the reads per water sample in Sainte Anne and La Claye, respectively, on average). This pattern could be the result of the capacity of some rozellids (e.g. *Rozella* spp.)

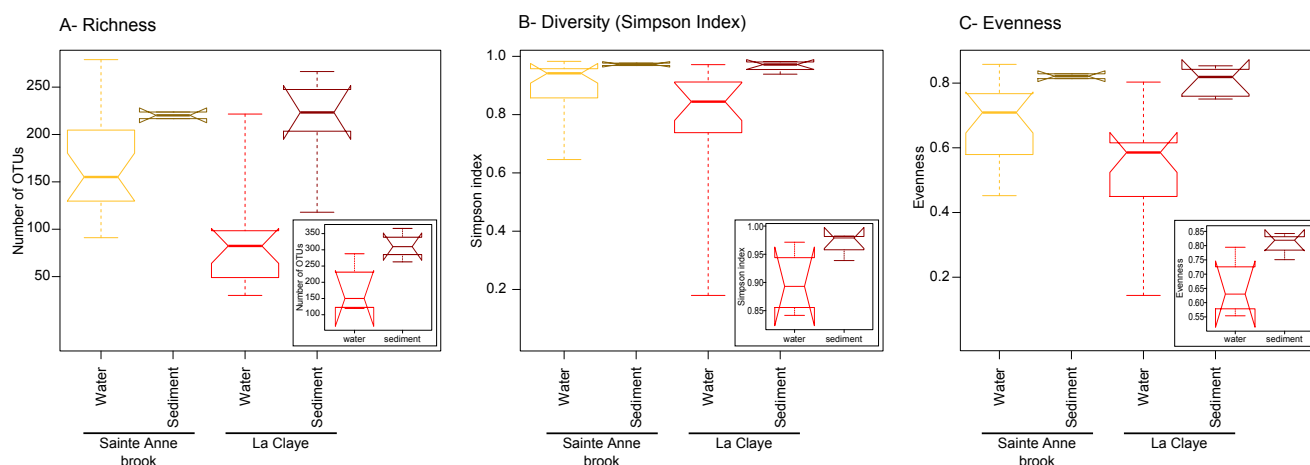
to produce resting sedimenting forms (Lara *et al.* 2010) or reflect the distribution of potential rozellids' hosts. Indeed, cryptomycota cells have been observed attached to diatoms (Jones *et al.* 2011) and were recently described infecting amoebozoan hosts (Corsaro *et al.* 2014), both taxa being more abundant in sediment. Indeed, Amoebozoans were always recorded at significant abundances in the sediments (2.1% of the reads per sample, on average) while they were merely detectable in water habitat (0.03% of the reads per sample, on average). Cercozoans too reached higher abundances in the sediment than in the water in La Claye (10.4% and 2.9% of the reads per sediment and water samples respectively, on average) (Figure 1 and 2A). This group was reported at significant proportion in soil (Moon-van der Staay *et al.* 2006) and sediments (Takishita *et al.* 2007) and at variable abundance over time and depth in Arctic meromictic lake (Charvet *et al.* 2012). Within the taxonomic groups recorded at higher abundance in the water habitat, Chrysophyceae-Synurophyceae was the main group in Sainte Anne brook planktonic communities (27.2% of the reads per sample, on average) while their mean relative abundance in the sediment fell by nearly 13 times (2.1% of the reads per sample, on average), and were also usually more abundant in La Claye sediment but their abundance could be occasionally quite low in water samples too (Figure 1). Cryptophytes are usual and quantitatively important members of planktonic communities in freshwaters (Lefranc *et al.*, 2005; Šlapeta *et al.*, 2005; Taib *et al.*, 2013). They were the main (26.5% of reads per sample on average) and fourth (11.1%) more abundant group in water habitat of La Claye pond and Sainte Anne brook respectively. Yet, they were detected at very low abundances in the sediment of these two systems (0.03% and 0.8% of the reads per sample, respectively) (Figure 2A), as reported for an oxic pond (Šlapeta *et al.*, 2005). Despite those major differences, other taxa were globally encountered in slightly different abundances in sediment and planktonic samples. Fungi were globally more abundant in sediment than in water in both systems, but they could occasionally reach high relative abundance within water communities where they have a marked dynamics (Figure 1). Streptophytes were in general more abundant in sediment than in water communities in La Claye pond but they occasionally reached high relative proportions in planktonic communities too (Figure 1), suggesting that the difference we observe for that group could partially be due to temporal dynamics. Glaucophytes, which are rarely detected in molecular environmental surveys (Massana *et al.* 2004a; Countway *et al.* 2007) (e.g. Massana *et al.*, 2004; Countway *et al.*, 2007), were only recorded in La Claye pond sediments. They were represented by two OTUs. OTU 2491 (99% identical to sequence X81901, *Gloeochaete wittrockiana*) and OTU 2122 (99% identical to sequence X68483, *Cyanophora paradoxa*) occurred at very low abundance (2 reads each) in

sediment samples in October and November 2011, respectively. Interestingly, OTU 2122 was also detected two times in planktonic communities of another shallow pond in the same geographical region (Simon *et al.*, in prep, see Chapter 3), suggesting that this phylotype might not be restricted to the sediment habitat. Interestingly, Dictyochophyceae, which is rarely detected in aquatic environments (e.g. Kirkham *et al.*, 2011), was only detected in water here, where it represented up to 5.3% of the reads per sample in La Claye pond. Perkinseans, are typical of freshwaters (e.g. Richards *et al.*, 2005; Lepère *et al.*, 2008), and yet, they were slightly more abundant within the water than in the sediment communities (Figure 2A). Sediment samples were only collected in summer and autumn while planktonic samples cover the four seasons. Yet, the main differences observed here between sediment and water communities (i.e. more amoebozoans, rostellids, cercozoans and diatoms, and less cryptophytes and chrysophyceae in the sediment) are likely not triggered by seasonality as they were also detected when comparing the same number of sediment and water eukaryote communities of the same period in La Claye pond (Figure 2B). The differences of taxonomic composition of the eukaryote assemblages between the two habitats (Figures 1 and 2) could also be influenced by the prefiltrations of water samples designed to select cells in the 0.2-5  $\mu\text{m}$  size-fraction while no size selection was applied on sediment communities. Even if that bias may be real for the biggest and more unbending cells, it must be limited according to our previous observation that assemblages in the 0.2-5  $\mu\text{m}$  or in the 5-30  $\mu\text{m}$  size-fraction obtained from planktonic samples in another shallow pond by successive filtrations harbored highly similar diversities (Simon *et al.*, accepted).

Differences in taxonomic community compositions were concomitant to slight differences in the functional composition of the eukaryote communities in La Claye pond. Indeed, putative phototrophic organisms generally dominated planktonic communities while no clear dominance of any functional group was observed for the sediment communities (Figure S1). In addition, putative parasites, especially rostellids, were more abundant in the sediment than in the water habitats, when comparing either all samples or the same number of water and sediment samples of the same period of the year. All these results thus suggest that droughts cause drastic shift in the taxonomic composition of the eukaryote community inhabiting small and shallow ecosystems.

Then, sediment and water communities differed in their structure, the former being richer and more diverse than the latter, as indicated by the distribution of richness, Simpson and evenness indices differing for water and sediment communities (Figure 4). We also

compared eukaryote communities in the same number of sediment and water samples one year apart but at the same period in La Claye pond to avoid differences due to different seasons. The comparison, based on fewer samples but less biased, confirmed the differences of richness and diversity detected when analysing complete datasets of water and sediment eukaryote communities.



**Figure 4: Differences in richness and diversity between all water and sediment communities in Sainte Anne brook and La Claye pond.** In the insets are shown the distributions of richness and diversity indices among four sediment samples in summer and autumn 2011 and four water samples of the same period in 2012. Richness is defined as the expected number of OTUs in a subsample of the same size as the smallest sample included in the study. The thick line represents the median of the distribution; the lower and upper limits of the boxes correspond to the first and third quartile respectively. Whiskers extend to the minimal and maximal values. Notches are drawn to indicate whether medians from distinct distributions can be considered as different.

In addition, the eukaryote communities of the same ecosystem were more similar in the sediment than in the water (Figures 3 and S2), with mean pairwise Bray-Curtis dissimilarities of 0.85 and 0.68 between water and sediment communities of the same ecosystem, respectively. In La Claye pond, the mean pairwise dissimilarity values between the same number of water or sediment samples, from the same period of the year, were of 0.85 and 0.65 respectively. These results suggest that the higher variability observed for the water communities was not due to the larger sampling period in this habitat. Planktonic communities might have thus stronger temporal dynamics than their sediment counterparts, but comparisons on longer surveys are needed.

Finally, sediment communities from Sainte Anne brook and La Claye pond were more similar one to each other than planktonic assemblages, based on pairwise Bray-Curtis dissimilarities (Figures 3 and S3). Sediment eukaryote assemblages might thus display a lowest spatial variability than water communities, at least at this restricted geographical scale.

Sediment and planktonic communities may differ in their proportion of resting cells as inactive taxa can be twice more abundant in soil than in freshwaters (Lennon & Jones 2011). Also, a large part of the OTUs detected in the sediment may correspond to dormant cells, free DNA or dead cells coming from planktonic communities that sedimented before or during the drought period. These OTUs could maybe partially explain the largest richness and diversity observed in the sediment assemblages as well as their lower spatio-temporal variability.

Altogether, communities retrieved in the water and the sediment showed clear different signatures.

#### *Recovery of planktonic communities after droughts.*

Such a drastic event as total drying up of small and shallow ecosystems constitutes an important stress and may thus trigger drastic changes in the aquatic communities even after the systems are flooded again, which, in turn, may affect the functioning of the ecosystem. Although there was a methodological difference in the measure of the eukaryote community diversity in water and sediments, i.e. with and without size-fractionation of the sample, no biases affected the analysis of the planktonic community recovery after drought events as we compared planktonic communities before and after drought, analysed in the same way.

We observed only slight changes in the small eukaryote composition after a severe drought of ca. 5 months in La Claye pond, from late July 2011 to early January 2012 (Figure 3B) and no community composition shift after a drought of 1-2 months in both La Claye pond and in Sainte Anne brook (Figure 3B-C). Small planktonic eukaryotes thus seem to be able to recover after short and long drought events but more data are needed to confirm these observations and better appreciate the possible long-term impact of long droughts on this community.

At a shorter time-scale, the first water sample after La Claye was filled up again in January 2012 (LC19) was typical of planktonic eukaryotes (Figure 3). On the contrary the water eukaryote communities from the first water samples after the shorter second drought event in summer 2012 (LC37 and RSA37) were closer to sediment communities than any other planktonic communities (Figure 3B-C). Communities sampled one month later (i.e. from samples LC39 and RSA39, November 2012) presented once again a planktonic

signature. One possible explanation for that difference between the two drought periods could be the lapse of time between the rain episodes and the sampling. Indeed, the two major rain episodes following the first drought occurred 12 to 7 days and most importantly 29 to 22 days before sampling LC19 (January 2012) while they took place 19 to 13 and above all 7 to 1 days before sampling LC37 and RSA37 (October 2012) at the end of the second drought (Figure S4). So, communities from surface water may look like sediment samples during a few days after the ecosystems are filled up again, but return to a typical planktonic signature rapidly.

Overall, our observations suggest that planktonic eukaryotes have the capacity to recover to severe droughts within a month once the ecosystem has been filled again with water. Such a rapid recovery after drastic changes in environmental conditions could be surprising as it was not observed for sediment bacterial assemblages in a semi-permanent stream that was subjected to drought (Rees *et al.* 2006). However, the rapid resilience of planktonic microbial communities has already been reported for other disturbances such as drastic water column mixing (Shade *et al.* 2012b).

Aquatic communities are exposed to important and rapid seasonal variations in the water physico-chemical characteristics in small and shallow freshwater ecosystems and are highly dynamics (Bamforth, 1958; Simon *et al.*, in prep, see Chapter 3) as for Sainte Anne brook and La Claye pond. In addition, short and long droughts can occur regularly and seasonally in small freshwater systems, driving rapid shifts of the water flows and from wet to dry conditions, as it happened in the brook and pond we studied. Macro-organism assemblages undergoing regular seasonal drought are characterized by a strong recovery capacity, likely because of the selection of organisms that use refuge or can adopt a life-stage that can stand desiccation (Lake 2003). Microbial communities inhabiting particular environments such as ponds and brooks might also encompass taxa adapted to desiccation. Indeed, we identified a few OTUs that were very resistant to drought events in our two systems. For instance, OTU 290 (93% identical to DQ104595, *Rozella* sp.) and OTU 356 (99% identical to DQ244023, *Stichotrichia*) were detected during the whole 2-years survey, including sediment and planktonic assemblages, in Sainte Anne brook and La Claye pond, respectively. The planktonic eukaryote community in Sainte Anne Brook and La Claye pond may thus be partly composed of taxa capable of dormancy life stages that wake up when water is back. Immigration from geographically close and non-disturbed ecosystems surrounding La Claye pond or from upstream parts of the brook is another potential

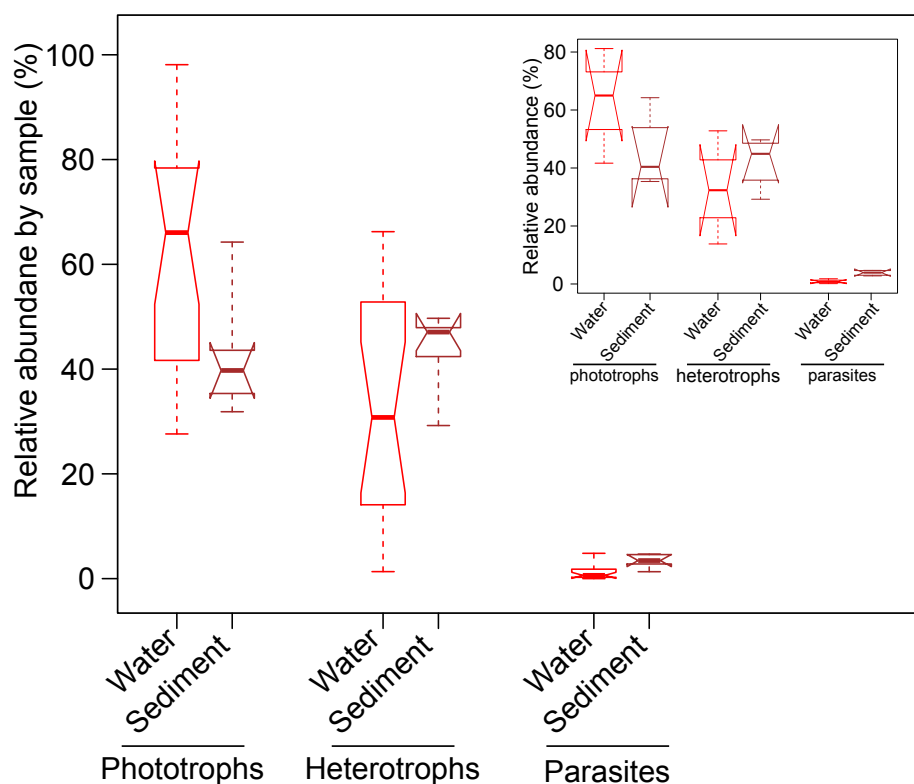


explanation for microbial recovery (Shade *et al.* 2012a) and could allow sustaining taxa that are not able to survive in dry sediment conditions in small and shallow freshwater ecosystems.

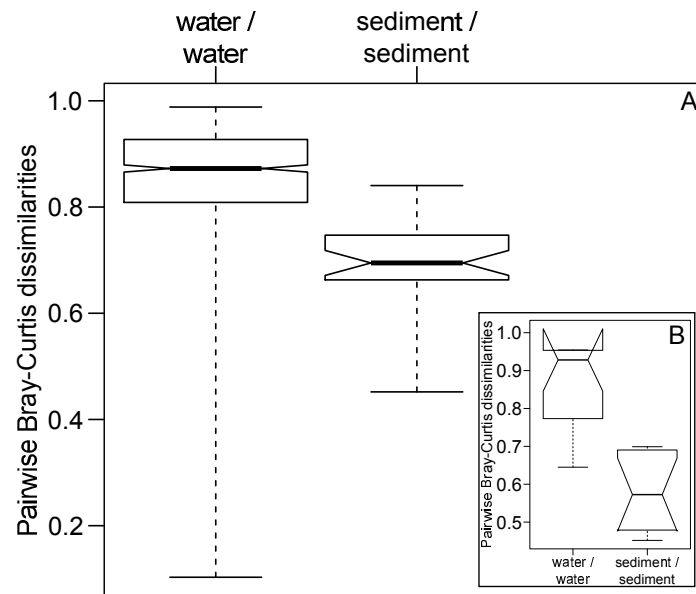
### E- Conclusions

Eukaryote communities in sediment and water habitats in one brook and one pond were characterised by marked differences in their taxonomic composition, structure and spatio-temporal dynamics. Eukaryotic assemblages detected in the sediment appeared richer and more diverse but less variable in time and space than their planktonic counterparts. Planktonic communities recovered within a month after drought periods. These observations reveal that planktonic eukaryote communities of small and shallow freshwater ecosystems harbor a high resilience. Sampling at higher frequency, daily or weekly, will help clarifying the time needed to recover and deciphering the microbial successions leading to the recovery of this community after drought events.

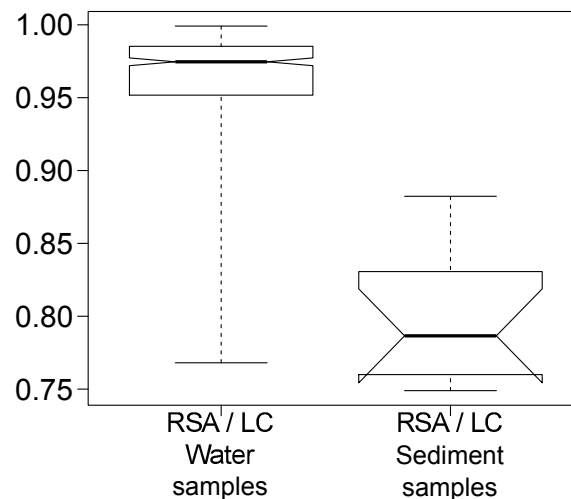
## F- Supplementary Information



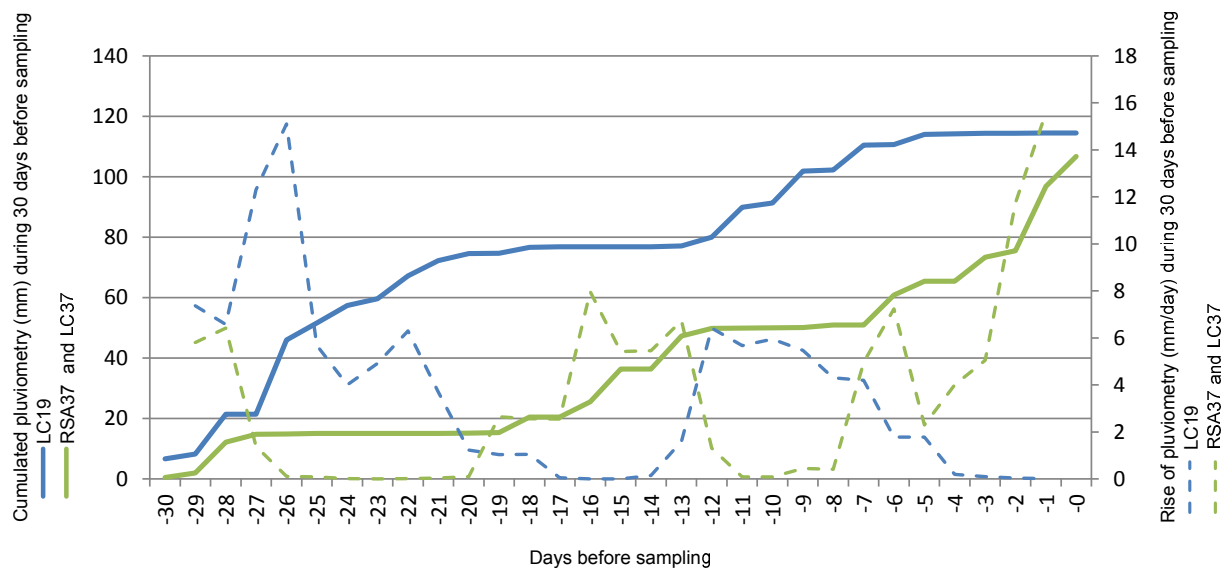
**Figure S1:** Distribution of the relative abundance of putative phototrophs, parasites and other heterotrophs in water and sediment communities in La Claye pond. Large graph: all samples in La Claye, Inset: four sediment samples from summer and autumn 2011 and four water samples from the same period in 2012. Thick lines indicate median values. The top and below limits of boxes indicate the third and first quartile, respectively. Whiskers extend to the minimal and maximal values. Non-overlapping notches are strong evidence that medians differ.



**Figure S2:** Distribution of pairwise Bray-Curtis dissimilarities conducted on OTU percentages, between all (A) samples from the same ecosystem and of the same type (water or sample), or (B) between four sediment samples from summer / autumn 2011 and between four water samples from the same period in 2012. Thick horizontal bars represent the median of the distributions; the lower and upper borders of the boxes correspond to the first and third quartiles, respectively. Whiskers extend to the minimal and maximal values. Non-overlapping notches are strong evidence that medians differ.



**Figure S3:** Distribution of pairwise Bray-Curtis dissimilarities between water or sediment communities from distinct ecosystems. RSA: Sainte Anne brook, LC: La Claye pond. Thick lines correspond to the medians of the distributions. Non-overlapping notches are strong evidence that medians differ. Lower and upper limits of the boxes indicate the first and third quartiles, respectively. Whiskers extend to the minimal and maximal values of distributions. Non-overlapping notches are strong evidence that medians differ.



*Figure S4: Pluviometry during the month before the first samplings following rewetting. Rise of pluviometry at day  $d$  is estimated as half the difference between pluviometry at day  $d+1$  and at  $d-1$ .*



# Quelques éléments de discussion générale

---



*"Les petites choses ont leur importance ; c'est toujours par elles qu'on se perd."*  
Fiodor Dostoïevski, *Crime et châtiment*





# Quelques éléments de discussion générale

---

## I. La richesse inexplorée des mares et ruisseaux

Il me semble quelque peu étonnant que la diversité microbienne dans les petits systèmes d'eau douce ait été si peu étudiée par des méthodes moléculaires jusqu'ici. Malgré les intérêts évidents que présente ce type de systèmes, évoqués en introduction et au cours du manuscrit, malgré l'observation par les méthodes traditionnelles d'une diversité non négligeable de micro-organismes eucaryotes et de leur dynamique, il est surprenant que les mares et petits ruisseaux aient été généralement exclus des études de plus en plus nombreuses et dans des environnements divers et variés. Les écosystèmes dans lesquels la diversité des eucaryotes microbiens a été étudiée sont souvent des environnements considérés extrêmes ou en tout cas plus exotiques que la mare d'à côté (e.g. haute mer, océan profond, milieux hypersalés). Peut-être mares et ruisseaux sont-ils victimes de l'impression commune que tout est plus intéressant loin de chez soi, et que ce que l'on voit tous les jours n'a rien d'intéressant à révéler. Pourtant, malgré leur apparente banalité, les petits écosystèmes d'eau douce peu profonds pourraient héberger une part encore inconnue de la diversité eucaryote.

Explorer la diversité des petits eucaryotes (fraction de taille théorique 0,2-5  $\mu\text{m}$ ) nous a permis de détecter, dans quatre mares et un ruisseau, une grande diversité d'organismes, dans tous les supergroupes reconnus et dans certains taxa dont la position dans l'arbre des eucaryotes n'est pas clairement définie. Les cryptophytes, les straménopiles (essentiellement Chrysophyceae-Synurophyceae), les alvéolés (essentiellement ciliés) et les champignons constituaient les groupes les plus abondants sur l'ensemble des sites et sur les deux ans de suivi. De ce point de vue, les communautés de micro-organismes eucaryotes de nos mares ressemblent à celles des grands lacs (e.g. Richards, 2005; Taib et al., 2013) ou des rares mares précédemment étudiées (Šlapeta *et al.* 2005).

A une échelle plus fine, certaines OTU dans nos bibliothèques de séquences branchaient profondément au sein de différents supergroupes. Elles ont notamment pour 1<sup>er</sup> hit par blast des séquences affiliées aux colpodellides, groupe frère des apicomplexés au sein des alvéolés (Leander *et al.* 2003), aux opisthosporides, groupe frère des champignons au sein des opisthocontes (Karpov *et al.* 2014) ou aux Nouveaux Clades 10 et 12 et Tremulida au sein des cercozoaires (Bass *et al.* 2009; Howe *et al.* 2011). Ces lignées branchant à la base de groupes

sont particulièrement intéressantes, en particulier parce qu'elles peuvent aider à reconstruire l'histoire évolutive de caractères au sein des groupes considérés (Howe *et al.* 2011), du moins si on connaît leur phénotype. De nombreuses autres OTU étaient affiliées à des taxa dont la position phylogénétique au sein de l'arbre des eucaryotes n'est pas résolue. Certaines étaient ainsi affiliées aux groupes Haptophyta, Cryptophyta, Telonemida, Katablepharida, Centroheliozoa, Apusozoa. L'augmentation de l'échantillonnage taxonomique étant une des façons d'améliorer les phylogénies, la diversité au sein de ces groupes détectable dans les milieux d'eau douce peu profonds pourrait aider à déterminer leur position dans l'arbre des eucaryotes.

Ensuite, les petits écosystèmes d'eau douce peu profonds semblent héberger des organismes dont il serait particulièrement intéressant de connaître la biologie. C'est le cas notamment du groupe EV, groupe de prymnésiphytes (Haptophytes) découvert dans l'étang des Vallées. Les OTU de ce groupe sont affiliées aux Isochrysidales, un ordre majoritairement composé d'organismes formant des écailles calcifiées (coccolithophores), alors que jusqu'ici un seul coccolithophore a été décrit en eau douce, *Hymenomonas roseola*, affilié d'après sa morphologie aux Coccolithales (Manton *et al.* 1969).

Ainsi, le petit échantillon d'écosystèmes d'eau douce peu profonds étudié ici révèle la présence dans ce type de milieux d'une très grande diversité de micro-organismes eucaryotes, dont des organismes particulièrement intéressants de par leur position phylogénétique ou les questions qui se posent quant à leur biologie. Comme nous l'avons évoqué en introduction, les mares et ruisseaux sont extrêmement nombreux et hétérogènes à la surface du globe. Il est donc plus que probable que l'exploration de la diversité des eucaryotes microbiens dans une plus large gamme de ces écosystèmes permettrait de découvrir toute une part encore inconnue de la diversité de ce domaine. Caractériser cette diversité permettrait de mieux comprendre le fonctionnement écologique de ces écosystèmes importants à l'échelle locale et mondiale (Downing *et al.* 2006; Downing 2010), et fournirait des données utiles pour tenter de résoudre certaines relations phylogénétiques au sein de l'arbre des eucaryotes.

En outre, les petits écosystèmes d'eau douce peu profonds constituent un environnement de choix pour étudier différents aspects de l'écologie des micro-organismes eucaryotes en général. Comme nous l'avons évoqué, ces systèmes semblent particulièrement adaptés à l'étude de l'impact relatif de la distance et des conditions environnementales locales sur la répartition observable des micro-organismes, puisqu'ils constituent un milieu structuré, discontinu, et sont extrêmement diversifiés. D'autre part, les mares représentent également un cadre de choix pour l'analyse de la dynamique temporelle des eucaryotes microbiens et des

micro-organismes en général. Tout d'abord, ces petits volumes d'eau stagnants ne sont pas soumis à de grands courants comme en milieu marin : les différences observées dans la composition de communautés prélevées au même endroit à deux moments différents ont plus de chances de correspondre à l'évolution temporelle d'une même communauté, et non au remplacement de la première communauté par une autre, déplacée depuis un peu plus loin par le courant. Ensuite, de nombreux ruisseaux et mares sont aisément accessibles, certains à proximité des laboratoires car très largement distribués, ce qui permet la mise en place simple de suivis à long terme ou un échantillonnage fréquent.

Etudier plus avant les petits écosystèmes d'eau douce peu profonds enrichirait donc certainement notre connaissance des micro-organismes en général et des eucaryotes en particulier.

## II. La barrière entre milieu marin et eau douce, moins infranchissable que prévu?

La frontière entre milieux marin et d'eau douce, et notamment la grande différence de salinité de l'un à l'autre, semble constituer la barrière la plus difficile à franchir par les macro-organismes, et constitue donc une forte limitation à leur dispersion (Logares *et al.* 2009). Chez les bactéries également, il a été montré que les différences de salinité représentaient le principal facteur environnemental corrélé avec les différences de composition des communautés (Lozupone & Knight 2007).

Différentes observations semblent indiquer que la frontière entre milieu marin et eau douce est également difficilement franchissable par les micro-organismes eucaryotes. Tout d'abord, certains groupes très abondants et répandus en milieu marin n'ont jamais été observés en eau douce. C'est le cas notamment des Syndiniales (=MALV I et II, alvéolés, (López-García *et al.* 2001; Guillou *et al.* 2008), ou de la plupart des groupes MAST (straménopiles, Massana *et al.*, 2013), qui d'après les connaissances actuelles semblent ne jamais avoir colonisé les eaux douces, suggérant que les transitions d'un milieu à l'autre puissent être particulièrement difficiles dans certains groupes. Au sein des taxa dont on connaît des ressortissants en milieu marin et en eau douce, les analyses moléculaires montrent que les séquences d'ADNr 18S se regroupent généralement dans les arbres phylogénétiques en fonction du type de milieu à partir duquel elles ont été obtenues, indépendamment de leur origine géographique (ce qui, notons le, est en faveur de l'hypothèse de Baas-Becking). Ces clusters de séquences issues d'environnements d'eau douce, distincts des séquences d'origine marine, ont été observés dans différents taxa, dont les foraminifères, les choanoflagellés

## Discussion générale

(Logares *et al.* 2009), des straménopiles hétérotrophes (Richards *et al.* 2005), les perkinsozoaires (Bråte *et al.* 2010b), ou les télonémides (Bråte *et al.* 2010a) par exemple. Tous les organismes d'eau douce d'un même cluster au sein d'organismes marins sont supposés découler d'un même événement de colonisation de l'eau douce, suivi par une diversification dans ce nouveau milieu. Ainsi, le fait que les organismes trouvés en eau douce se regroupent généralement ensemble, en un nombre limité de clusters, est interprété comme la preuve d'un faible nombre de transitions du milieu marin vers l'eau douce (Logares *et al.* 2009). De plus, la relativement grande diversité dans de nombreux clades d'eau douce a suggéré que la plupart des transitions soient anciennes (Logares *et al.* 2009). Nous avons fait au cours de ce travail certaines observations similaires. En effet, la diversité d'haptophytes détectée en eau douce est largement inférieure à celle détectée en milieu marin. Notamment, nous avons détecté une bien plus grande variété de phylotypes affiliés au prymnésiophytes en milieu marin qu'en eau douce, alors que plus d'échantillons d'eau douce, de beaucoup plus de sites, ont été explorés, et que les échantillons ont été analysés exactement de la même manière dans les deux types de milieu (chapitre 1). En considérant les données issues de notre étude basée sur le clonage / séquençage Sanger des ADNr 18S et du suivi basé sur le pyroséquençage 454 de ce même marqueur, ainsi que celles présentes dans la littérature (en particulier Šlapeta *et al.*, 2005; Shalchian-Tabrizi *et al.*, 2011), les haptophytes pourraient avoir colonisé les eaux douces avec succès au cours de dix transitions indépendantes au moins. Ce nombre peut sembler relativement faible compte tenu de l'âge des haptophytes dont la racine est datée à plus de 800 millions d'années (Liu *et al.* 2010), et les grandes diversité et abondance actuelles et passées au sein de ce groupe. On notera que si la position de la lignée HAP-1 (Šlapeta *et al.* 2005; Shalchian-Tabrizi *et al.* 2011) branchant à la base de l'arbre des haptophytes est confirmée, les transitions vers l'eau douce ont commencé très tôt dans l'histoire du groupe. Cela pourrait même remettre en question l'origine marine des haptophytes. De la même manière que pour les haptophytes, la plupart des pyroséquences affiliées aux groupes MAST (chapitre 2) se regroupent dans nos reconstructions phylogénétiques en clusters de séquences issues d'échantillons d'eau douce. En particulier, plus des trois quart de ces pyroséquences appartiennent au groupe MAST-12, et majoritairement à MAST-12C, subclade dans lequel nombre de séquences d'eau douce sont déjà connues (Massana *et al.* 2013). Ainsi, on ne rencontre généralement pas les mêmes organismes en eau douce et en milieu marin, ce qui suppose que le passage d'un milieu à l'autre n'est pas anodin.

Cette observation s'explique assez aisément : alors que la plupart des milieux marins ont une salinité telle que milieu extérieur et cytoplasme ont des concentrations en osmolites

relativement similaires, l'eau douce est fortement hypotonique. Les cellules dans l'eau douce sont donc en permanence en déséquilibre osmotique, et doivent lutter activement contre l'entrée d'eau, faute de quoi elles gonfleraient jusqu'à ce que la membrane plasmique rompe, notamment dans le cas des cellules sans paroi. Outre le risque d'éclatement, l'entrée d'eau et la sortie d'ions selon le gradient de potentiel osmotique peuvent influencer le métabolisme cellulaire. On rencontre différentes adaptations en eau douce liées à la régulation du volume cellulaire. En particulier, les vacuoles pulsatiles sont rencontrées chez de nombreux eucaryotes unicellulaires phylogénétiquement distants ; elles permettent, en se contractant de façon cyclique, d'expulser une partie du cytosol pour compenser l'entrée continue d'eau dans la cellule (Hoef-Emden 2014). On peut donc supposer que passer du milieu marin proche de l'isotonie à l'eau douce très hypotonique nécessite la mise en place d'un certain nombre d'adaptations, qui rendent difficiles les transitions d'un milieu à l'autre, d'autant plus que les nouveaux arrivants se retrouvent en compétition avec des organismes déjà adaptés (Logares *et al.* 2009).

Cependant, diverses OTU obtenues au cours de ce travail de thèse à partir d'échantillons d'eau douce ne se regroupaient pas en clusters distincts des séquences d'origine marine. Par exemple, l'OTU 3828 détectée dans le ru Sainte Anne en avril et mai 2012 par pyroséquençage présente plus de 99% d'identité avec la séquence de l'ADNr 18S d'*Ochrosphaera verrucosa* (Haptophytes), espèce marine, et branche avec la séquence CPT2-Pry1-C2 détectée dans un écosystème continental salé. De même, l'OTU 152 affiliée au groupe MAST-2 présente 99% d'identité avec des séquences environnementales issues d'échantillons d'eau douce mais aussi marins. Cette grande proximité entre séquences d'origine marine et d'eau douce dans MAST-2 avait déjà été observée et interprétée comme la conséquence d'une colonisation très récente des eaux douces au sein de ce groupe (Massana *et al.* 2013). De même, au sein des dinoflagellés, *Peridinium aciculiferum*, rencontrée en eau douce, et *Scrippsiella hangoi*, marine, ont des ITS aux séquences identiques (Logares *et al.* 2009). Ainsi, même si de nombreuses transitions milieu marin / eau douce semblent anciennes (Logares *et al.* 2009), différents indices laissent à penser que des transitions ont également eu lieu récemment, au sein de différents groupes.

En plus de pouvoir être plus récentes que prévu, ces transitions pourraient également être plus fréquentes qu'initialement supposé. Nous avons ainsi pu détecter dans le ru Sainte Anne deux OTU affiliées au clade MAST-3 (Stramenopiles), jusqu'ici considéré comme exclusivement marin, ainsi que diverses OTU affiliées par BLAST à divers groupes MAST, mais dont l'affiliation exacte reste à définir. De même, Alverson et collègues (2007) ont



## Discussion générale

montré l'existence de quatre colonisations indépendantes au minimum des eaux douces depuis les océans rien qu'au sein de la famille des Thalassiosirales (Bacillariophyceae, Stramenopiles). Les colonisations du milieu marin depuis l'eau douce étaient supposées encore plus rares (Alverson *et al.* 2007), mais différentes études ont mis en évidence des transitions dans ce sens, par exemple au sein des perkinsozoaires (Bråte *et al.* 2010b) ou des diatomées (Alverson *et al.* 2007). Ainsi, les transitions entre le milieu marin et l'eau douce sont sans doute plus fréquentes que supposé, malgré la grande différence de salinité donc d'osmolarité de l'un à l'autre.

Dans les petits écosystèmes d'eau douce peu profonds, dont le volume d'eau varie fortement au cours du temps, on peut s'attendre à des variations non négligeables de l'osmolarité. En particulier, la conductivité dans La Claye et le ru Sainte Anne variait entre 251 et 533  $\mu\text{S}/\text{cm}$ , et 0,6 et 1 247  $\mu\text{S}/\text{cm}$  respectivement. En outre, lors de l'assèchement total puis la remise en eau de ces écosystèmes, les organismes présents ont sans doute dû faire face à des chocs osmotiques importants. Pourtant, on a retrouvé des OTU présentes dans un grand nombre voire dans tous les échantillons prélevés au cours des deux ans de suivi dans ces écosystèmes, ce qui suppose l'existence d'organismes tolérant des variations importantes de l'osmolarité de leur milieu de vie. Cette tolérance a été montrée au sein de plusieurs groupes, et associée à la présence de vacuoles pulsatiles actives même en milieu marin (Hoef-Emden 2014). Kaneshiro et collègues (1969) ont ainsi montré que des ciliés marins pouvaient être cultivés dans de l'eau de mer diluée à 50%. Lorsqu'ils sont changés de milieu, leur volume cellulaire augmente, puis rediminue suite à l'augmentation du débit d'expulsion des vacuoles pulsatiles. De même, certains cryptophytes marins comme *Chroomonas placodea* paraissent tolérer une large gamme d'osmolarité, grâce à l'augmentation du volume et de la fréquence de contraction des vacuoles pulsatiles (Hoef-Emden 2014). Cette large tolérance peut permettre aux organismes concernés de rapidement s'adapter à leur nouveau milieu, malgré une forte différence de salinité.

Ainsi, le passage marin / eau douce constitue une barrière à la dispersion des micro-organismes eucaryotes, sans doute liée aux différences de salinité donc d'osmolarité ; on observe généralement des organismes différents dans les deux types de milieu. Cependant, le nombre de transitions de l'un à l'autre pourrait être beaucoup plus élevé que ce que proposaient Logares et collègues (2009), du moins dans certains taxa. Un plus vaste échantillonnage d'écosystèmes d'eau douce plus variés révélera sans doute que la barrière entre marin et eau douce a été franchie encore plus souvent que ce que l'on voit actuellement.

### III. Dynamique spatio-temporelle des communautés de petits eucaryotes dans les mares et ru

Comme évoqué en introduction, les petits écosystèmes d'eau douce peu profonds présentent des caractéristiques environnementales extrêmement diversifiées, même au sein d'une même région géographique (Oertli & Frossard 2013). Bien qu'une analyse conjointe des paramètres mesurés (ACP) n'ait pas permis de distinguer les cinq écosystèmes que nous avons sélectionnés, ils représentent bien cette grande diversité, de par leur environnement proche (en forêt, dans un hameau, au bord d'une route, dans un complexe de mares), leur dynamique (eau stagnante ou courante), leur taille (quelques centaines à quelques milliers de m<sup>2</sup>) et divers paramètres physico-chimiques (notamment pH, conductivité, concentrations en NO<sub>3</sub><sup>-</sup>, NH<sub>3</sub>, et chlorophylle *a*). De façon intéressante, les communautés de petits eucaryotes étaient très différentes d'un écosystème à l'autre, et même plus différents que d'une saison à l'autre. Les différences entre les assemblages prélevés en avril 2012 n'étaient pas corrélées avec les distances géographiques entre sites. De même, tout au long du suivi nous avons pu observer la présence en forte abondance de certains phylotypes dans des écosystèmes donnés sans que l'on ait pu les détecter dans le système le plus proche. En revanche, en avril 2012, les différences entre communautés étaient significativement corrélées aux différences environnementales mesurées dans les différents sites. Ainsi, à la petite échelle spatiale que nous considérons ici (distance maximale entre 2 écosystèmes < 10 km), on n'observe pas de structuration par la distance des communautés de petits eucaryotes des mares et rus, mais des assemblages différents en fonction des caractéristiques environnementales. Les communautés étudiées semblent donc suivre ici l'assertion de Baas-Becking "*Everything is everywhere, but, the environment selects*" (de Wit & Bouvier 2006).

Les petits écosystèmes d'eau douce peu profonds sont également caractérisés par des variations importantes et rapides de leurs caractéristiques environnementales, comme illustré en introduction (Angélibert *et al.* 2004). Les communautés qui y vivent sont ainsi soumises en permanence à des changements de leur environnement en fonction des saisons et / ou d'événements plus ponctuels. Quelques rares suivis réalisés dans des écosystèmes d'eau douce peu profonds ont d'ailleurs montré que la composition de ces communautés variait de façon importante au fil du temps (Bamforth 1958; Lara *et al.* 2011). Au sein de chacun des écosystèmes étudiés, une forte dynamique temporelle a pu être observée à différents niveaux.

Tout d'abord, les communautés de petits eucaryotes étaient très variables d'un mois à l'autre, avec les médianes des indices de dissimilarité de Bray-Curtis entre échantillons

## Discussion générale

successifs variant de 0,61 (Etang des Vallées) à 0,79 (Saint Robert et La Claye). Cette dynamique rapide se traduisait par des pics d'abondance relative très élevée et relativement courts (un ou deux mois la plupart du temps) de différents groupes taxonomiques. Ces pics étaient soit réguliers, avec des alternances rapides de forte et plus faible abondance relative (e.g. Cryptophytes), soit plus ponctuels (e.g. Katablépharides dans La Claye et Saint Robert). A l'échelle des phylotypes, certaines OTU étaient régulièrement détectées et participaient à cette rapide dynamique par les variations de leur abondance relative essentiellement, alors que d'autres apparaissaient et disparaissaient, une ou plusieurs fois, de nos *pools* de séquences. Il est difficile de savoir à quoi correspondent ces détections ponctuelles. Certaines des OTU concernées pourraient être toujours présentes dans l'écosystème, mais être détectées seulement occasionnellement. Dans ce cas, la détection dans certains échantillons plutôt que dans d'autres pourrait être due aux effets stochastiques. En effet, beaucoup des OTU rarement détectées le sont dans des abondances relatives très faibles. Il est possible alors que l'on puisse souvent et aléatoirement ne pas les détecter au cours d'une des étapes, parce que les organismes correspondant n'étaient pas dans le volume d'eau filtré, leur ADN n'était pas dans le volume introduit dans la réaction de PCR, ou bien parce que leur ADNr 18S était moins facilement amplifié ou séquencé que celui d'un organisme abondant à ce moment-là. Alternativement, leur détection à un moment donné plutôt qu'à un autre pourrait être dû à une augmentation non négligeable de leur abondance relative, grâce à laquelle elles passent au-dessus des limites de détection de la méthode utilisée, et ainsi traduire la dynamique temporelle de ces phylotypes. C'est sans doute le cas au moins pour les OTU détectées ponctuellement mais avec une abondance élevée. Une autre partie des OTU détectées à un moment donné et jamais avant pourrait également correspondre à des organismes nouvellement arrivés dans l'écosystème, depuis d'autres sites. La quantité d'immigrants pourrait ne pas être négligeable étant données la densité de milieux humides dans la région, et l'absence de différenciation par la distance des communautés. Leur disparition des banques de séquences ultérieures pourrait être due à leur disparition de l'écosystème ou à une forte diminution de leur abondance relative. Toutes ces possibilités coexistent probablement, mais leur importance relative n'est pas connue.

Au cours des deux ans d'échantillonnage, les communautés présentes dans La Claye et dans le ru Sainte Anne ont subi, en plus des variations habituelles de leur environnement, des perturbations fortes. Le débit de Sainte Anne était très variable, le ru s'est résumé pendant de courtes périodes à une succession de flaques non connectées. Il a traversé une période d'assèchement total entre août et septembre 2012. De même, La Claye était totalement à sec lors des échantillonnages de juillet à décembre 2011 puis en septembre 2012. Ces périodes de

sécheresse imposaient certainement un stress important aux communautés planctoniques (e.g. stress hydrique et stress osmotique associé, perte de leur milieu de locomotion, de nutrition). Pourtant, les communautés récupéraient en moins d'un mois, montrant une grande stabilité.

L'existence d'une saisonnalité dans la dynamique des micro-organismes et notamment des eucaryotes est une question toujours en suspens. Plusieurs études indépendantes ont montré une augmentation de l'abondance de groupes données aux mêmes saisons en différents endroits (Cf. introduction générale), suggérant que certains groupes au moins ont une saisonnalité. Cependant, les suivis réalisés sur plus d'un an ne montrent pas tous de dynamique répétable d'une année à l'autre (Lepère *et al.* 2006; Jones *et al.* 2013; Kim *et al.* 2014). Aussi, des événements ponctuels ou des différences climatiques d'une année à l'autre pourraient parfois être plus importantes que les variations environnementales saisonnières. Les petits écosystèmes d'eau douce peu profonds sont particulièrement sensibles à la fois aux variations saisonnières de leur environnement (Angélibert *et al.* 2004) et aux perturbations ponctuelles (e.g. augmentation de la concentration en nutriments par épandage d'engrais aux alentours, pollution accidentelle, période exceptionnellement pluvieuse). Ils semblent donc particulièrement adaptés à l'étude de la part relative des facteurs saisonniers et des événements occasionnels. Ici, malgré les événements perturbateurs importants que sont les périodes d'assèchement total, nous avons pu détecter une dynamique saisonnière marquée des communautés de petits eucaryotes dans nos mares et ru. En effet, les assemblages échantillonnés dans des saisons opposées étaient sensiblement plus différents que des assemblages échantillonnés au cours de la même saison. Notamment, les communautés d'été et d'hiver étaient très différentes, alors que celles d'automne et de printemps étaient plus intermédiaires, certains échantillons étant plus proches de ceux d'été, et d'autres de ceux d'hiver. Le graphique montrant la moyenne des dissimilarités de Bray-Curtis entre échantillons séparés du même laps de temps indiquait une saisonnalité moins claire pour La Claye, mais ceci peut être dû à l'absence de 6 échantillons (pendant la sécheresse) : les moyennes étant calculées sur un nombre plus restreint d'indices de dissimilarité sont certainement moins fiables que pour les autres écosystèmes. Nous avons cherché à savoir quels organismes étaient responsables de cette saisonnalité, en regardant la dynamique des taxa à un rang taxonomique relativement élevé, ainsi qu'au niveau des phylotypes, et n'avons pu détecter de saisonnalité claire qu'au sein de ces derniers. Les taxa considérés regroupaient peut être une trop grande diversité écologique pour pouvoir distinguer des préférences saisonnières communes aux organismes de ce groupe. Peut-être une saisonnalité pourrait-elle être observée pour des taxa de rang taxonomique inférieur, comme ça a été observé pour des ordres de bactéries (Gilbert *et al.* 2012). Plusieurs OTU montraient ainsi une dynamique

## Discussion générale

saisonnaire marquée, avec des pics d'abondance observés les deux années à la même période, comme l'OTU 2986 (Bicosoécides, Straménopiles) en hiver ou l'OTU 15 (Raphidophycées, Straménopiles) en été. Dans tous les cas, sans doute grâce à leur résilience après une perturbation ponctuelle, les assemblages de petits eucaryotes semblent être majoritairement influencés par les variations saisonnières de leur environnement : variations de la température, de l'intensité et de la durée de l'ensoleillement quotidien, et des paramètres biotiques qui en découlent (e.g. abondance des bactéries autotrophes, activité des macrophytes). Un suivi à plus long terme sera nécessaire pour vérifier la régularité de cette observation.

Différentes études se sont attelées à corréler paramètres environnementaux et dynamique des communautés de micro-organismes eucaryotes en milieu aquatique, et ne parvenaient généralement à en expliquer qu'une partie (Lepère *et al.* 2006; Kirkham *et al.* 2011). Dans notre étude, la dynamique spatio-temporelle sur 2 ans des assemblages de petits eucaryotes n'était que partiellement expliquée par les paramètres environnementaux, physico-chimiques et climatiques, pris en compte. En effet, on a estimé que seuls 22% de la variabilité des 37 OTU présentes dans au moins 25% des 112 échantillons de plancton de la fraction de taille théorique 0,2-5  $\mu\text{m}$ , et représentant au moins 0.1% des séquences en moyenne dans ces mêmes échantillons, étaient expliqués par l'ensemble des données environnementales considérées. Ceci peut s'expliquer en partie par le fait que nous n'avons pris en compte aucune interaction biotique, même si les interactions hôte-pathogène, entre symbiotes mutualistes ou proie-prédateurs sont potentiellement déterminantes pour expliquer la structure des communautés eucaryotes. De même, les virus (Cf. Introduction générale) pourraient également être responsables d'une partie de la dynamique que nous n'avons pas pu expliquer ici. Cependant, les paramètres potentiellement impliqués dans la dynamique des micro-organismes étant tellement nombreux, les interactions entre tous si complexes, et les caractéristiques environnementales étant potentiellement si variables à l'échelle micro-locale, il est extrêmement difficile d'expliquer entièrement la dynamique d'une communauté de micro-organismes avec les données actuellement disponibles.

Cependant, une analyse multivariée réalisée sur la dynamique des 37 OTU citées ci-dessus au cours des 2 ans et sur l'ensemble des paramètres physico-chimiques et climatiques considérés nous a permis de mettre en évidence différentes corrélations plus ou moins prononcées entre des groupes taxonomiques donnés et certains paramètres environnementaux. La corrélation la plus claire associe les OTU de champignons aux hautes valeurs de conductivité. Comme toute corrélation, elle est à interpréter avec précautions. L'interprétation la plus séduisante consisterait à voir là un lien de cause à effet, les hautes valeurs de

conductivité favorisant la croissance des champignons, explication très envisageable, compte tenu notamment de leur mode de vie par osmotrophie. Mais il pourrait également s'agir d'une corrélation sans lien de causalité puisque hautes valeurs de conductivité et forte abondance des champignons se rencontrent toutes deux dans le ru Sainte Anne, et pourraient toutes deux être les conséquences, indépendantes l'une de l'autre, d'une contamination des échantillons d'eau de ce très petit système par du sédiment contenant des spores. D'autres corrélations, moins nettes, ont pu être suspectées entre chlorophytes et concentrations élevées en phosphate, entre cryptophytes et valeurs relativement élevées de la teneur en oxygène dissous, de la concentration en nitrite ou du pH, ou entre ciliés et faible conductivité. Ces observations devront être vérifiées sur un plus grand nombre d'OTU au sein de chaque taxon, mais il est intéressant de noter que l'on peut trouver des corrélations communes à des phylotypes variés au sein d'un grand groupe taxonomique. Cependant, on remarquera également que les différentes OTU affiliées à un même groupe peuvent être associées à des valeurs très différentes de certains paramètres. Il semble ainsi que l'on puisse trouver des chlorophytes à toutes conductivités, ou des ciliés indépendamment des concentrations en chlorophylle et nitrite, de la teneur en oxygène ou du pH, par exemple. Il apparaît donc particulièrement important de prendre en compte la variabilité intra-groupe lorsque l'on souhaite identifier des corrélations propres à l'ensemble du groupe. Considérer l'abondance totale d'un taxon sans regarder les différences à une échelle plus fine risquerait en effet de mettre en évidence des corrélations concernant le ou les phylotypes les plus abondants, qui pourraient être confondues avec des caractéristiques du taxon dans son entier même si elles ne sont pas du tout partagées par certaines lignées peut être moins abondantes au sein du taxon.

#### IV. Complémentarité du clonage / séquençage Sanger et du pyroséquençage 454

Les premières études de la diversité microbienne basées sur le séquençage de fragments de l'ADNr 18S obtenaient les séquences par séquençage Sanger des amplicons séparés par clonage en bactéries (Cf. Introduction pour le principe de la méthode). Depuis 2004 jusqu'à ces dernières années, cette méthode a été la plus utilisée pour déterminer la diversité microbienne (Zinger *et al.* 2011). Les étapes de PCR impliquées dans cette méthodes sont susceptibles de produire des amplicons ne correspondant à aucun gène d'ADNr 18S réellement existant, via la création de chimères (recombinaison entre deux fragments de gènes d'organismes différents), ou l'introduction de mutations (Zinger *et al.* 2011). En revanche, l'étape de séquençage en elle-même, par les développements actuels de la méthode de Sanger,



## Discussion générale

est fiable, produisant des séquences avec 99,999% d'exactitude par base (Shendure & Ji 2008). De plus, les séquences obtenues lors d'une réaction de séquençage sont longues (environ 1000 pb, Shendure & Ji, 2008), et la possibilité de séquencer chaque amplicon à partir de chacune de ses extrémités permet l'obtention de séquences plus longues encore, couvrant une grande partie de l'ADNr 18S. Cette méthode nécessite cependant des manipulations relativement longues : PCR, clonage, repiquage d'un grand nombre de clones, PCR sur chacun de ces clones pour réamplifier les amplicons isolés, éventuellement purification des produits PCR. Le temps et le matériel nécessaires à la réalisation de ces manipulations augmente avec le nombre de clones étudiés, donc avec le nombre de séquences que l'on veut obtenir. De plus, le séquençage Sanger d'un amplicon est relativement cher, et séquencer un grand nombre de fragments d'ADN peut prendre beaucoup de temps. Aussi les études utilisant le clonage et séquençage Sanger des amplicons sont-elles basées sur un nombre relativement limité de séquences, généralement quelques centaines par échantillon au grand maximum (López-García *et al.* 2001; Lefranc *et al.* 2005; Richards *et al.* 2005).

Avec la volonté croissante d'explorer en profondeur la diversité des micro-organismes dans de nombreux écosystèmes, de pouvoir séquencer de plus en plus de génomes complets ou d'utiliser en routine le séquençage comme outil de diagnostic médical, la nécessité d'une méthode de séquençage plus rapide et moins couteuse se fait de plus en plus pressante. Plusieurs méthodes de séquençage ne nécessitant plus le passage par une étape de clonage et basées sur des principes variés voient le jour dans le courant des années 2000. Le pyroséquençage 454 est la première de ces technologies de séquençage nouvelle génération (NGS) à avoir été mise sur le marché (Margulies *et al.* 2005; Shendure & Ji 2008). C'est également celle qui est actuellement la plus utilisée pour caractériser la diversité microbienne (Zinger *et al.* 2011), notamment parce que jusqu'à récemment c'était la NGS produisant des séquences plus longues (entre 300 et 500 pb environ, Gilles *et al.*, 2011). Cette méthode permet de séquencer en une seule fois les amplicons issus de plusieurs dizaines d'échantillons (jusqu'à 20 échantillons par quart de *run*). La technologie généralement utilisée actuellement (454 GS-FLX Titanium, Roche) permet de produire environ un million de séquences en un *run* de 10 h (Gilles *et al.* 2011). De plus, les NGS en général et le pyroséquençage 454 en particulier permettent de diminuer drastiquement le coût par base séquencée. Ainsi, ces nouvelles technologies permettent de séquencer rapidement, à moindre coût, des amplicons issus de nombreux échantillons, grâce à quoi on peut facilement comparer la diversité présente à différentes dates ou en différents sites. La grande profondeur de séquençage permise par ces nouvelles technologies permet en outre de détecter les amplicons les plus rares issus de la PCR sur l'ADN total extrait de l'échantillon d'intérêt. C'est pour ces deux

raisons que nous avons choisi de réaliser le suivi mensuel sur deux ans de la diversité générale des micro-organismes eucaryotes par pyroséquençage 454 (technologie GS-FLX Titanium, Roche). En seulement deux *runs* de séquençage, nous avons pu séquencer les amplicons correspondant aux échantillons de 5 écosystèmes pendant 24 mois ainsi que 24 échantillons d'une autre fraction de taille pour l'un des écosystèmes et 8 duplicats, soit 152 échantillons au total. Les coûts financier et temporel qu'aurait engendrés un tel échantillonnage si l'on avait utilisé clonage et séquençage Sanger auraient été inabordables. En outre, nous avons pu détecter, grâce à la grande profondeur de séquençage du 454, des OTU extrêmement peu abondantes, qui ne semblent pas être dues à des artefacts de manipulations puisqu'elles ont été observées dans des échantillons indépendants et/ou ont des séquences très proches dans les bases de données. Par exemple, des deux OTU affiliées aux Glaucophytes (Archaeplastida) détectées, l'une n'est représentée que par deux séquences, l'autre (OTU 2122, 99,6% d'identité avec une séquence de *Cyanophora paradoxa*) est détectée dans trois échantillons. L'organisme correspondant à cette OTU semble donc être effectivement un membre des communautés de petits eucaryotes des mares, mais elle ne représentait jamais plus de 0.1% des séquences de l'échantillon où on l'a détectée. Elle n'aurait donc probablement pas été détectée par clonage et séquençage Sanger *via* lesquels on obtient bien moins d'un millier de séquences par échantillon. De la même manière, les OTU de centrohéliozaïres et d'excavés représentent respectivement moins de 0,3% des séquences dans 9 des 11 et 12 des 16 échantillons où on les a détectées. La grande profondeur du pyroséquençage 454 nous a donc permis de détecter la présence et de mettre en évidence la relativement grande fréquence de différents taxa, ce que nous aurions pu ne pas voir en utilisant clonage et séquençage Sanger.

Pour ces mêmes raisons, les NGS suscitent un engouement croissant de la part des microbiologistes cherchant à caractériser la diversité microbienne. Au point qu'il est aujourd'hui difficile de publier une analyse de diversité basée uniquement sur la méthode classique par clonage et séquençage Sanger. Pour autant, les NGS sont-elles toujours la méthode la plus appropriée?

Nous avons étudié la diversité au sein du groupe des haptophytes dans divers écosystèmes d'eau douce, à la fois par clonage et séquençage Sanger d'amplicons obtenus à l'aide d'amorces spécifiques du groupe, et par pyroséquençage 454 de produits PCR amplifiés grâce à des amorces générales d'eucaryotes. Malgré la grande profondeur du pyroséquençage, cette technique ne nous a pas permis de détecter une plus large diversité que l'amplification à l'aide d'amorces spécifiques du groupe de fragments d'ADNr 18S suivie par leur clonage et séquençage Sanger. La grande majorité des OTU détectées par pyroséquençage étaient proches de celles obtenues par clonage / séquençage Sanger. En outre, cette seconde méthode

## Discussion générale

nous a permis de détecter une plus grande diversité au sein du nouveau groupe d'isochrysidales de l'étang des Vallées. Ainsi, dans ce cas précis d'exploration d'un groupe relativement peu divers dans l'environnement étudié, le pyroséquençage et sa grande profondeur ne sont pas nécessaires.

D'autre part, les NGS, et notamment le pyroséquençage 454, ne sont pas parfaites. Outre les séquences artefactuelles dues aux erreurs lors de la PCR initiale - les mêmes que pour le clonage / séquençage Sanger - le séquençage de nouvelle génération en lui-même introduit de fréquentes erreurs. Gilles et collègues (2011) ont montré que presque 90% des séquences complètes produites par la technologie 454 GS-FLX Titanium (Roche) et ayant passé les filtres de qualité du séquenceur contiennent au moins une erreur, ce qui correspond à plus de 0,9% d'erreurs par bases contre 0.001% pour le séquençage Sanger (Shendure & Ji 2008). De plus, les séquences produites, même par le pyroséquençage 454, sont courtes (autour de 400 pb), ce qui constitue une faible quantité d'informations lorsque l'on souhaite les intégrer à des phylogénies. Il est notamment difficile de déterminer de façon à peu près fiable la position phylogénétique de séquences courtes sans séquence proche dans les bases de données, et l'on peut difficilement assurer être en présence d'un groupe inconnu sur la base de séquences issues des NGS. Dans notre cas, la diversité et l'affiliation du nouveau groupe d'haptophytes EV ont pu être mises en évidence grâce aux séquences longues issues du clonage / séquençage Sanger. Ainsi, les NGS ne sont non seulement pas toujours nécessaires mais donnent également des résultats parfois moins adaptés aux questions posées que les méthodes moléculaires plus traditionnelles.

Aussi, comme les caractérisations morphologiques restent indispensables malgré la nécessité indéniable d'utiliser des approches moléculaires, il me semble que la technique moléculaire classique basée sur le clonage et le séquençage ne doit pas être délaissée pour les études de diversité malgré l'attrait de la nouveauté et les avantages des nouvelles technologies de séquençage, du moins tant que celles-ci n'ont pas atteint le même niveau de précision et la même longueur de séquence que le Sanger. Les NGS comme le pyroséquençage 454 sont extrêmement utiles quand il s'agit de décrire qualitativement et *semi*-quantitativement des communautés diverses, et de comparer les assemblages microbiens entre différents sites et dates. En revanche, lorsque l'on souhaite avoir des informations plus précises sur un groupe donné, déterminer les relations phylogénétiques entre différents taxa, le clonage / séquençage Sanger semble plus adapté, notamment après avoir ciblé les ADNr 18S du groupe d'intérêt à l'aide d'amorces *spécifiques* pour en détecter la plus grande diversité possible. Les deux méthodes peuvent être complémentaires, la seconde permettant de préciser des éléments particulièrement intéressants mis en évidence par la première.





# Conclusion

---

## I. Les principales informations de ce travail

Ce travail de thèse visait principalement à décrire la diversité des micro-organismes eucaryotes dans des écosystèmes peu explorés que sont les mares et ruisseaux. Nous souhaitions également avoir un premier aperçu de leur écologie dans ces systèmes. Dans ces objectifs, nous avons exploré la diversité intra-groupe au sein des haptophytes, un taxon méconnu en eau douce, dans un large panel d'écosystèmes, grâce à l'amplification à l'aide d'amorces spécifiques du groupe d'un fragment d'ADNr 18S, suivie du clonage et séquençage Sanger des amplicons obtenus. Par la suite, nous avons suivi la composition des communautés de petits eucaryotes en général dans quatre mares et un ru, par pyroséquençage 454 de fragments d'ADNr 18S amplifiés à partir d'échantillons prélevés mensuellement durant deux ans. Ces travaux nous ont notamment permis de montrer :

- Une très grande diversité des micro-organismes eucaryotes dans des écosystèmes d'eau douce peu profonds, couvrant l'ensemble des supergroupes connus ainsi que différents taxa dont la position dans l'arbre est méconnue.
- L'existence d'un groupe de séquences de prymnésiophytes (haptophytes) proches des isochrysodales, dont la branche principale est très divergente. Ce groupe a pu être détecté dans deux écosystèmes distants et différents, le lac d'Annecy et l'étang des Vallées. Il montre une forte saisonnalité, étant détectable dans l'étang des Vallées tous les ans aux mêmes périodes, entre 2010 et 2013.
- La présence en eau douce d'une variété d'organismes proches d'eucaryotes observés en milieu marin, au sein de différents taxa. Notamment, nous avons détecté diverses OTU affiliées plus ou moins clairement à des groupes de MArine STramenopiles (MAST), dont le groupe MAST-3 jusqu'ici considéré comme exclusivement marin. Ces observations suggèrent l'existence de transition entre milieu marin et eau douce plus récentes et plus fréquentes que prévu, du moins dans certains groupes phylogénétiques.
- Une grande dynamique des communautés de petits eucaryotes. La composition des communautés (plus que leur richesse ou diversité) présente une composante



## Conclusion

saisonnière marquée, due à la forte saisonnalité de certaines OTU. Ces communautés font preuve d'une forte résilience et se reconstituent rapidement ( $< 1$  mois) après l'assèchement total de leur écosystème, événement perturbateur fort.

- ④ La présence dans des écosystèmes géographiquement proches de communautés très différentes. Malgré leur importante variabilité temporelle, elles conservent une signature propre à chaque écosystème. Ces différences inter-écosystèmes sont corrélées avec les différences de caractéristiques physico-chimiques des mares et ru, et non avec la distance géographique séparant les sites. A cette échelle très locale de distances géographiques, il semble donc que les conditions environnementales locales soient la principale cause de la répartition géographique observable. Cependant, nous n'avons pas pu corréler de façon indiscutable un paramètre précis avec la répartition géographique de taxa donnés, sauf pour les champignons particulièrement abondants quand la conductivité est élevée.
- ④ Les limites du fractionnement des échantillons par filtrations en fonction du diamètre des cellules. Les communautés sensées correspondre aux organismes de  $0,2$  à  $5\ \mu\text{m}$  et de  $5$  à  $30\ \mu\text{m}$  sont très semblables, notamment au niveau des OTU majoritaires.

## II. Quelques perspectives

Au-delà des résultats évoqués ci-dessus, les travaux présentés dans ce manuscrit appellent quelques approfondissements et ouvrent la voie à d'autres questions.

***Décrire la biologie des organismes correspondant au groupe de séquences d'haptophytes divergentes de l'étang des Vallées.***

Enfin, nous avons mis en évidence dans l'étang des Vallées l'existence d'un groupe de séquences particulièrement intéressant, parce que sa branche principale est divergente par rapport aux séquences connues, qu'il branche au sein des coccolithophores alors que très peu d'espèces portant des coccolithes sont décrites en eau douce, et qu'il présente une saisonnalité très marquée. Il serait donc particulièrement intéressant de mettre un visage sur ces séquences, de pouvoir observer la morphologie (notamment savoir s'ils possèdent ou non des écailles calcifiées), le métabolisme et l'écologie des organismes associés à ces séquences. En parallèle des échantillonnages de la première année de suivi, nous avons mis en culture des échantillons de chaque écosystème dans différents milieux. La présence des organismes

correspondant au groupe de séquences d'intérêt dans certaines de ces cultures pourra être testée par PCR et séquençage Sanger. S'ils sont effectivement présents, il sera possible d'essayer de dessiner des sondes spécifiques de ce groupe pour les observer par FISH, ainsi que de tenter de les isoler, par dilutions sériées par exemple, pour les observer vivants.

***Explorer les paramètres biotiques influençant la structure des communautés de micro-organismes eucaryotes.***

Tout d'abord, au cours de cette thèse, nous n'avons pris en compte que les paramètres physico-chimiques et météorologiques susceptibles d'être impliqués dans la répartition géographique et la variabilité temporelle des micro-organismes eucaryotes détectés. Or, les interactions biotiques avec les autres membres de la communauté d'un écosystème jouent sans doute un rôle essentiel dans la dynamique spatio-temporelle des assemblages microbiens. Aussi, au cours de l'échantillonnage, nous avons également collecté des échantillons du plancton animal, dont les organismes pourront par la suite être dénombrés et identifiés afin de voir d'éventuelles corrélations entre la dynamique des prédateurs et celle de leurs proies. Des fragments d'ADNr 16S de bactéries et archées ont également été amplifiés à partir des mêmes échantillons d'ADN que pour le suivi des communautés eucaryotes, puis pyroséquencés. L'analyse conjointe de ces pools de séquences et de celles des ARNr 18S d'eucaryotes pourrait permettre d'observer d'éventuelles corrélations entre les dynamiques d'eucaryotes bactérovores et de proies. D'autre part, la diversité des bactéries et archées dans les petits écosystèmes d'eau douce peu profonds n'ayant pas non plus été beaucoup étudiée, l'exploration de cette diversité pourrait être particulièrement intéressante, et donner une vue plus intégrée des biocénoses de ces systèmes.

***Décrire précisément le comportement des communautés de micro-organismes eucaryotes lors des périodes d'assèchement total de l'écosystème, voire lors d'autres événements perturbateurs.***

Ensuite, nous avons eu l'occasion au cours du suivi temporel de décrire l'évolution des communautés au cours d'épisodes d'assèchement total de la mare La Claye et du ru Sainte Anne. Ces observations nous ont permis de constater une résilience rapide des communautés planctoniques après la remise en eau. Cependant, ces épisodes n'étant pas prévus, l'échantillonnage n'était pas adapté à une description précise de l'événement et beaucoup de points restent à tester ou au moins à vérifier. Il serait particulièrement intéressant de réaliser un suivi des communautés plus fréquent (tous les jours ou tous les 2 à 3 jours), particulièrement pendant la période d'assèchement et celle de remise en eau. Ceci permettrait

## *Conclusion*

en particulier d'avoir une idée plus précise de la vitesse à laquelle les communautés prélevées dans l'eau retrouvent une signature de communauté planctonique, et par quelles étapes passe la récupération. Il pourrait également être intéressant d'analyser les communautés présentes dans les sédiments pendant que l'écosystème est toujours en eau, afin de pouvoir comparer les assemblages de micro-organismes eucaryotes de l'eau et des sédiments sur un même nombre d'échantillons prélevés simultanément, afin de limiter les risques de confondre dynamique temporelle et différences plancton / benthos.

### ***Suivre la dynamique temporelle des communautés d'eucaryotes microbiens à plus long terme.***

Nous avons d'autre part constaté des indices de saisonnalité dans la dynamique des assemblages d'eucaryotes microbiens, au cours des deux ans de suivi. Un suivi plus long permettrait de confirmer, ou non, cette saisonnalité. Actuellement, les mêmes cinq écosystèmes continuent à être échantillonnés dans les mêmes conditions une fois par saison. Ceci pourra peut-être aussi permettre d'appréhender la dynamique des communautés à plus long terme, parallèlement au vieillissement des mares. Ce type d'informations pourrait éventuellement intéresser le Parc Naturel Régional, qui suit particulièrement l'évolution de ses mares depuis quelques années, ayant fait le constat d'une diminution de leur nombre sur son territoire.

Au-delà de ces approfondissements qui permettraient de compléter ou consolider les résultats déjà obtenus, d'autres questions pourraient être posées. Entre autres :

### ***Suivre les communautés de micro-organismes eucaryotes au cours de cycles jour-nuit.***

Comme évoqué en introduction, les petits écosystèmes d'eau douce peu profonds tels que les mares voient leurs caractéristiques physico-chimiques (e.g. température, pH, concentration en dioxygène dissous) varier fortement au cours d'une journée. Il pourrait donc être intéressant de regarder si, au cours d'un cycle jour / nuit, les communautés de micro-organismes eucaryotes de surface varient, par exemple par des migrations d'organismes vers et depuis le fond de la mare. On pourrait pour cela envisager un échantillonnage toutes les deux ou trois heures par exemple pendant quelques jours, en surface et plus près du sédiment.

**Caractériser génétiquement *Gonyostomum semen* (Raphidophyceae, Stramenopiles).**

Enfin, nous avons détecté dans la mare Gabard, chaque été, un organisme que je trouve particulièrement séduisant, notamment par sa capacité à dominer totalement les assemblages d'eucaryotes planctoniques. *Gonyostomum semen* (Raphidophyceae, Straménopiles) est détectée dans de nombreuses mares légèrement acides et très colorées à travers l'Europe, aux USA ou au Japon par exemple. Il s'agirait en fait de la Raphidophyceae la plus répandue dans le monde (Rengefors *et al.* 2012). Or, il n'y a actuellement que deux séquences, partielles, d'ADNr 18S de *Gonyostomum semen* référencées comme telles sur NCBI, dont une longue de moins de 450 paires de bases. Il s'agit donc essentiellement d'une morphoespèce, qui pourrait éventuellement cacher des espèces cryptiques. Il sera nécessaire d'obtenir des séquences d'ADNr et d'autres marqueurs génétiques d'organismes échantillonnés dans différentes régions du monde pour vérifier qu'il s'agisse bien de la même espèce et non de lignées distinctes morphologiquement semblables, comme c'est le cas par exemple pour *Micromonas pusilla* (Šlapeta *et al.* 2006). D'un point de vue pratique, ses efflorescences estivales récurrentes posent des problèmes écologiques, mais aussi économiques lorsqu'elles surviennent dans des écosystèmes d'agrément. Définir leur répartition au sein du territoire du parc pourrait donc éventuellement intéresser le PNR.



**Figure 1: Deux cellules de *Gonyostomum semen*, souche NIES-1380, isolée au Japon en eau douce. La séquence de l'ADNr 18S des cellules de cette souche présente 99,8% d'identité avec la séquence représentative de l'OTU 15 trouvée très dominante tous les étés dans la mare Gabard. (photo: <http://www.shigen.nig.ac.jp/algae/images/strainsimage/nies-1380.jpg>)**

# Bibliographie

---

- Adl SM, Simpson AGB, Lane CE *et al.* (2012) The revised classification of eukaryotes. *The Journal of eukaryotic microbiology*, **59**, 429–93.
- Alverson AJ, Jansen RK, Theriot EC (2007) Bridging the Rubicon: phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Molecular phylogenetics and evolution*, **45**, 193–210.
- Amaral Zettler LA, Gómez F, Zettler E *et al.* (2002) Eukaryotic diversity in Spain ' s River of Fire. *Nature*, **417**, 137.
- Amaral-Zettler LA, Rocca JD, Lamontagne MG, Dennett MR, Gast RJ (2008) Changes in microbial community structure in the wake of Hurricanes Katrina and Rita. *Environmental science & technology*, **42**, 9072–8.
- Andersen RA (2004) Biology and systematics of heterokont and haptophyte algae. *American journal of botany*, **91**, 1508–22.
- Andersen RA, Bidigare RR, Keller MD, Latasa M (1996) A comparison of HPLC pigment signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. *Deep Sea Research Part II: Topical Studies in Oceanography*, **43**, 517–537.
- Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.
- Angélibert S, Marty P, Céréghino R, Giani N (2004) Seasonal variations in the physical and chemical characteristics of ponds: implications for biodiversity conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **14**, 439–456.
- Baas-Becking L (1934) Geobiologie of inleiding tot de milieukunde. W.P. Van Stockum & Zoon, The Hague, the Netherlands.
- Bachy C, Dolan JR, López-García P, Deschamps P, Moreira D (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *The ISME journal*, **7**, 244–55.
- Bachy C, López-García P, Vereshchaka A, Moreira D (2011) Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the north pole at the end of the polar night. *Frontiers in microbiology*, **2**, 106.
- Baldauf SL (2008) An overview of the phylogeny and diversity of eukaryotes. *Journal of Systematics and Evolution*, **46**, 263–273.
- Bamforth SS (1958) Ecological Studies on the Planktonic Protozoa of a Small Artificial Pond. *Limnology and Oceanography*, **3**, 398–412.
- Barcat JA (2003) Robert Hooke (1635-1703). *Medicina*, **63**, 753–6.
- Bass D, Cavalier-Smith T (2004) Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *International journal of systematic and evolutionary microbiology*, **54**, 2393–404.

- Bass D, Chao EE-Y, Nikolaev S *et al.* (2009) Phylogeny of Novel Naked Filose and Reticulose Cercozoa: Granofilosea cl. n. and Proteomyxidea Revised. *Protist*, **160**, 75–109.
- Begon M, Townsend CR, Harper JL (2009) *Ecology: From Individuals to Ecosystems*. Blackwell Publishing.
- Behnke A, Barger KJ, Bunge J, Stoeck T (2010) Spatio-temporal variations in protistan communities along an O/HS gradient in the anoxic Framvaren Fjord (Norway). *FEMS microbiology ecology*, **72**, 89–102.
- Bendif EM, Probert I, Hervé A *et al.* (2011) Integrative taxonomy of the Pavlovophyceae (Haptophyta): a reassessment. *Protist*, **162**, 738–61.
- Berg LR, Raven PH, Hassenzahl DD. M (2009) *Environnement*. De Boeck Supérieur.
- Bertrand J-C, Caumette P, Lebaron P, Matheron R, Normand P (2011) *Ecologie microbienne. Microbiologie des milieux naturels et anthropisés*. Presses universitaires de Pau et des pays de l'Adour.
- Bolker B, Lodewijk B, Gentleman R *et al.* (2012) gplots: Various R programming tools for plotting data.
- Borcard D, Legendre P, Drapeau P (1992) Partialling out the Spatial Component of Ecological Variation. *Ecology*, **73**, 1045–1055.
- Bower SM, Carnegie RB, Goh B *et al.* (2004) Preferential PCR Amplification of Parasitic Protistan Small Subunit rDNA from Metazoan Tissues. *The Journal of Eukaryotic Microbiology*, **51**, 325–332.
- Bråte J, Klaveness D, Rygh T, Jakobsen KS, Shalchian-Tabrizi K (2010a) Telonemia-specific environmental 18S rDNA PCR reveals unknown diversity and multiple marine-freshwater colonizations. *BMC microbiology*, **10**, 168.
- Bråte J, Logares R, Berney C *et al.* (2010b) Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *The ISME journal*, **4**, 1144–53.
- Bray JR, Curtis JT (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, **27**, 325–349.
- Bromham L (2009) Why do species vary in their rate of molecular evolution? *Biology letters*, **5**, 401–4.
- Burki F, Inagaki Y, Bråte J *et al.* (2009) Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome biology and evolution*, **1**, 231–8.
- Burki F, Okamoto N, Pombert J-F, Keeling PJ (2012) The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proceedings. Biological sciences / The Royal Society*, **279**, 2246–54.
- Butcher RW (1952) Contributions to our knowledge of the smaller marine algae. *Journal of the Marine Biological Association of the United Kingdom*, **31**, 175.
- Caron DA (1994) Inorganic nutrients, bacteria, and the microbial loop. *Microbial ecology*, **28**, 295–8.

- Caron DA, Countway P (2009a) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Applied and Environmental Microbiology*, **75**, 5797–5808.
- Caron DA, Countway P (2009b) Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquatic Microbial Ecology*, **57**, 227–238.
- Caron D a, Countway PD, Jones AC, Kim DY, Schnetzer A (2012) Marine Protistan Diversity. *Annual Review of Marine Science*, **4**, 467–493.
- Carrias J-F, Amblard C, Quiblier-Lloberas C, Bourdier G (1998) Seasonal dynamics of free and attached heterotrophic nanoflagellates in an oligomesotrophic lake. *Freshwater Biology*, **39**, 91–101.
- Castresana J (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, **17**, 540–552.
- Céréghino R, Biggs J, Oertli B, Declerck S (2008) The ecology of European ponds: defining the characteristics of a neglected freshwater habitat. *Hydrobiologia*, **597**, 1–6.
- Chambers JM, Cleveland WS, Kleiner B, Tukey PA (1983) *Graphical methods for data analysis* (W&Brooks/Cole., Ed.).
- Chambouvet A, Morin P, Marie D, Guillou L (2008) Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science (New York, N.Y.)*, **322**, 1254–7.
- Charvet S, Vincent WF, Comeau A, Lovejoy C (2012) Pyrosequencing analysis of the protist communities in a High Arctic meromictic lake: DNA preservation and change. *Frontiers in microbiology*, **3**, 422.
- Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *The ISME journal*, **4**, 1053–9.
- Cho BC, Azam F (1988) Major role of bacteria in biogeochemical fluxes in the ocean's interior. *Nature*, **332**, 441–443.
- Cloern JE (1996) Phytoplankton bloom dynamics in coastal ecosystems: A review with some general lessons from sustained investigation of San Francisco Bay, California. *Reviews of Geophysics*, **34**, 127–168.
- Corliss JO (1984) The kingdom PROTISTA and its 45 phyla. *Biosystems*, **17**, 87–126.
- Corliss JO (1991) Historically important events, discoveries, and works in Protozoology from the mid-17th to the mid-20th century.
- Corliss JO (1998) Haeckel's kingdom protista and current concepts in systematic protistology. *Stapfia*, **56**, 85–104.
- Corliss JO (2002) A salute to Antony van Leeuwenhoek of Delft, most versatile 17th century founding father of protistology. *Protist*, **153**, 177–90.
- Corpe WA, Jensen TE (1992) An electron microscopic study of picoplanktonic organisms from a Small Lake. *Microbial ecology*, **24**, 181–97.



- Corsaro D, Walochnik J, Venditti D *et al.* (2014) Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. *Parasitology research*, **113**, 1909–18.
- Costanza R (1999) The ecological, economic, and social importance of the oceans. *Ecological Economics*, **31**, 199–213.
- Countway PD, Gast RJ, Dennett MR *et al.* (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). *Environmental microbiology*, **9**, 1219–32.
- Criscuolo A, Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC evolutionary biology*, **10**, 210.
- Cronberg G (2005) The life cycle of *Gonyostomum semen* (Raphidophyceae). *Phycologia*, **44**, 285–293.
- Cronberg G, Lindmark G, Björk S (1988) Mass development of the flagellate *Gonyostomum semen* (Raphidophyta) in Swedish forest lakes - an effect of acidification? *Hydrobiologia*, **161**, 217–236.
- Cuvelier M, Allen A, Monier A *et al.* (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proceedings of the ...*, **107**, 14679–84.
- Dawson SC, Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 8324–9.
- Derelle E, Ferraz C, Rombauts S *et al.* (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11647–52.
- Diez B, Pedrós-alió C, Massana R (2001) Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and Sequencing. *Applied and Environmental Microbiology*, **67**, 2932–2941.
- Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome research*, **15**, 330–40.
- Dobell C (1932) *Antony van Leeuwenhoek and his "Little Animals." Being Some Account of the Father of Protozoology and Bacteriology and his Multifarious Discoveries in these Disciplines. Collected, Translated, and Edited, from his Printed Works, Unpublished Manuscripts, a* (L John Bale, Sons & Danielsson, Ed.). New York, Harcourt, Brace and company.
- Dobson M, Frid C (2008) *Ecology of Aquatic Systems*. Oxford University Press, Oxford.
- Donaldson IM (2010) Robert Hooke's *Micrographia* of 1665 and 1667. *The journal of the Royal College of Physicians of Edinburgh*, **40**, 374–6.
- Downing JA (2010) Emerging global role of small lakes and ponds: little things mean a lot. *Limnetica*, **1**, 9–24.
- Downing JA, Duarte CM, Tranvik LJ *et al.* (2006) The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography*, **51**, 2388–2397.

- Dray S, Dufour A-B (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, **22**, 1–20.
- Duris P, Gohau G (2011) Chap.1 L'ordre de la création. In: *Histoire des sciences de la vie*
- Edwardsen B, Eikrem W, Green J (2000) Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. *Phycologia*, **39**, 19–35.
- Edwardsen B, Eikrem W, Throndsen J *et al.* (2011) Ribosomal DNA phylogenies and a morphological revision provide the basis for a revised taxonomy of the Prymnesiales (Haptophyta). *European Journal of Phycology*, **46**, 202–228.
- Edwardsen B, Imai I (2006) Ecology of Harmful Algae. In: *Ecological Studies*. (eds Granéli E, Turner JT). Springer Berlin Heidelberg.
- Edwardsen B, Medlin LK (2007) Unravelling the algae: the past, present, and future of algal systematics. In: *Unravelling the Algae: The past, Present and Future of Algal Systematics* (eds Brodie J, Lewis J), pp. 183–196. Oxford, UK: CRC Press, Taylor and Francis group.
- Fenchel T (1982) Ecology of Heterotrophic Microflagellates. I. Some Important Forms and Their Functional Morphology. *Marine Ecology Progress Series*, **8**, 211–223.
- Finlay B (2002) Global dispersal of free-living microbial eukaryote species. *Science*, **296**, 1061–1063.
- Foissner W (1997) Global soil ciliate (Protozoa, Ciliophora) diversity: a probability-based approach using large sample collections from Africa, Australia and Antarctica. *Biodiversity & Conservation*, **1638**, 1627–1638.
- Foissner W (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozoologica*.
- Fokin S (2013) Otto Bütschli (1848–1920): Where we will genuflect? *Protistology*.
- Foster RA, Collier JL, Carpenter EJ (2006) Reverse transcription PCR amplification of cyanobacterial symbiont 16S rRNA sequences from single non-photosynthetic eukaryotic marine planktonic host cells. *Journal of Phycology*, **42**, 243–250.
- Fracastoro G (1546) *De contagione et contagiosis morbis et eorum curatione*. Venise.
- Frada M, Probert I, Allen MJ, Wilson WH, De Vargas C (2008) The “Cheshire Cat” escape strategy of the coccolithophore *Emiliana huxleyi* in response to viral infection. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 15944–15949.
- Frias-Lopez J, Thompson A, Waldbauer J, Chisholm SW (2009) Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. *Environmental microbiology*, **11**, 512–25.
- Gajadhar AA, Marquardt WC, Hall R *et al.* (1991) Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Cryptosporidium parvum* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Molecular and biochemical parasitology*, **45**, 147–154.
- Garrison FH (1910) Fracastorius, Athanasius Kircher and the germ theory of disease. *Science (New York, N.Y.)*, **31**, 500–2.

- Gast RJ, McDonnell TA, Caron DA (2000) srDna-based taxonomic affinities of algal symbionts from a planktonic foraminifer and a solitary radiolarian. *Journal of Phycology*, **36**, 172–177.
- Gest H (2004a) The Discovery of Microorganisms Revisited. *ASM News-American Society for ....*
- Gest H (2004b) The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society. *Notes and Records of the Royal Society*, **58**, 187–201.
- Gilbert JA, Steele JA, Caporaso JG *et al.* (2012) Defining seasonal marine microbial community dynamics. *The ISME journal*, **6**, 298–308.
- Gilles A, Meglécz E, Pech N *et al.* (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC genomics*, **12**, 245.
- Giordan A (1987) *Histoire de la biologie*.
- Gleason FH, Carney LT, Lilje O, Glockling SL (2012) Ecological potentials of species of *Rozella* (Cryptomycota). *Fungal Ecology*, **5**, 651–656.
- Gómez F, Moreira D, Benzerara K, López-García P (2011) *Solenicola setigera* is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. *Environmental microbiology*, **13**, 193–202.
- Graham LE, Wilcox LW (2000) *Algae*.
- Green JC (1973) Studies in the fine structure and taxonomy of flagellates in the genus Pavlova . II. A freshwater representative, Pavlova granifera (Mack) comb. Nov. *British Phycological Journal*, **8**, 1–12.
- Green J, Bohannan BJM (2006) Spatial scaling of microbial biodiversity. *Trends in ecology & evolution*, **21**, 501–7.
- Green J, Holmes AJ, Westoby M *et al.* (2004) Spatial scaling of microbial eukaryote diversity. *Nature*, **432**, 747–50.
- Groissillier A, Massana R, Valentin K, Vaultot D, Guillou L (2006) Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquatic Microbial Ecology*, **42**, 277–291.
- Guillou L, Bachar D, Audic S *et al.* (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic acids research*, **41**, D597–604.
- Guillou L, Chrétiennot-Dinet M-J, Medlin LK *et al.* (1999) *Bolidomonas*: A new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). *Journal of Phycology*, **35**, 368–381.
- Guillou L, Viprey M, Chambouvet A *et al.* (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental microbiology*, **10**, 3349–65.
- Hansen LR, Kristiansen J, Rasmussen JV (1994) Potential toxicity of the freshwater Chrysochromulina species *C. parva* (Prymnesiophyceae). *Hydrobiologia*, **287**, 157–159.
- Harada A, Ohtsuka S, Horiguchi T (2007) Species of the Parasitic Genus *Duboscquella* are Members of the Enigmatic Marine Alveolate Group I. *Protist*, **158**, 337–347.

- Hartmann M, Grob C, Tarran GA *et al.* (2012) Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 5756–60.
- Heck KL, van Belle G, Simberloff D (1975) Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology*, **56**, 1459–1461.
- Hitchman RB, Jones HLJ (2000) The role of mixotrophic protists in the population dynamics of the microbial food web in a small artificial pond. *Freshwater Biology*, **43**, 231–241.
- Hoef-Emden K (2014) Osmotolerance in the Cryptophyceae: Jacks-of-all-trades in the Chroomonas Clade. *Protist*, **165**, 123–143.
- Hooke R (1665) Obferv. XX. Of blue Mould, and of the firft Principles of Vegetation arifing from Putrefaction. In: *Micrographia* , pp. 125–131.
- Howe AT, Bass D, Scoble JM *et al.* (2011) Novel Cultured Protists Identify Deep-branching Environmental DNA Clades of Cercozoa: New Genera *Tremula*, *Micrometopion*, *Minimassisteria*, *Nudifila*, *Peregrinia*. *Protist*, **162**, 332–372.
- Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577–586.
- IPCC (2007) Climate change 2007 : The Physical Science Basis. Contribution of working group 1. In: Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jardillier L, Zubkov M V, Pearman J, Scanlan DJ (2010) Significant CO<sub>2</sub> fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *The ISME journal*, **4**, 1180–92.
- Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC evolutionary biology*, **4**, 18.
- Johnson PW, Sieburth JM (1982) In-situ morphology and occurrence of eukaryotic phototrophs of bacterial size in the picoplankton of estuarine and oceanic waters. *Journal of Phycology*, **18**, 318–327.
- Jones MDM, Forn I, Gadelha C *et al.* (2011) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature*, **474**, 200–3.
- Jones AC, Liao TSV, Najar FZ *et al.* (2013) Seasonality and disturbance: annual pattern and response of the bacterial and microbial eukaryotic assemblages in a freshwater ecosystem. *Environmental microbiology*, **15**, 2557–72.
- Jones SE, McMahon KD (2009) Species-sorting may explain an apparent minimal effect of immigration on freshwater bacterial community dynamics. *Environmental microbiology*, **11**, 905–13.
- Jonkers HM, Ludwig R, Wit R *et al.* (2003) Structural and functional analysis of a microbial mat ecosystem from a unique permanent hypersaline inland lake: “La Salada de Chiprana” (NE Spain). *FEMS microbiology ecology*, **44**, 175–89.
- Jordan R, Chamberlain A (1997) Biodiversity among haptophyte algae. *Biodiversity & Conservation*, **6**, 131–152.

- Jordan R, Cros L, Young J (2004) A revised classification scheme for living haptophytes. *Micropaleontology*, **50**, 55–79.
- Kalff J, Knoechel R (1978) Phytoplankton and their dynamics in oligotrophic and eutrophic lakes. *Annual Review of Ecology and Systematics*, **9**, 475–95.
- Kamp A, de Beer D, Nitsch JL, Lavik G, Stief P (2011) Diatoms respire nitrate to survive dark and anoxic conditions. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 5649–54.
- Kaneshiro ES, Dunham PB, Holz GG (1969) Osmoregulation in a Marine Ciliate, *Miamiensis avidus*. I. Regulation of Inorganic Ions and Water. *Biological Bulletin*, **136**, 63.
- Karpov SA, Mamkaeva MA, Aleoshin V V. *et al.* (2014) Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Frontiers in Microbiology*, **5**.
- Katz LA (1998) Changing perspectives on the origin of eukaryotes. *Trends in Ecology & Evolution*, **13**, 493–497.
- Keeling PJ (2009) Chromalveolates and the evolution of plastids by secondary endosymbiosis. *The Journal of eukaryotic microbiology*, **56**, 1–8.
- Keller M, Selvin RC, Claus W, Guillard RRL (1987) Media for the culture of oceanic ultraphytoplankton. *Journal of Phycology*, **23**, 633–638.
- Kim DY, Countway PD, Jones AC *et al.* (2014) Monthly to interannual variability of microbial eukaryote assemblages at four depths in the eastern North Pacific. *The ISME journal*, **8**, 515–30.
- Kim E, Harrison J (2011) Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proceedings of the National Academy of Sciences*, **108**, 1496–500.
- Kirkham AR, Jardillier L, Tiganeşcu A *et al.* (2011) Basin-scale distribution patterns of photosynthetic picoeukaryotes along an Atlantic Meridional Transect. *Environmental Microbiology*, **13**, 975–990.
- Kirkham AR, Lepère C, Jardillier L *et al.* (2013) A global perspective on marine photosynthetic picoeukaryote community structure. *The ISME journal*, **7**, 922–36.
- Kolodziej K, Stoeck T (2007) Cellular identification of a novel uncultured marine stramenopile (MAST-12 Clade) small-subunit rRNA gene sequence from a norwegian estuary by use of fluorescence in situ hybridization-scanning electron microscopy. *Applied and environmental microbiology*, **73**, 2718–26.
- Kühn S, Medlin L, Eller G (2004) Phylogenetic position of the parasitoid nanoflagellate *Pirsonia* inferred from nuclear-encoded small subunit ribosomal DNA and a description of *Pseudopirsonia* n. gen. and *Pseudopirsonia mucosa* (Drebes) comb. nov. *Protist*, **155**, 143–56.
- Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental microbiology*, **12**, 118–23.
- Lake PS (2003) Ecological effects of perturbation by drought in flowing waters. *Freshwater Biology*, **48**, 1161–1172.

- Langenheder S, Jürgens K (2001) Regulation of bacterial biomass and community structure by metazoan and protozoan predation. *Limnology and oceanography*, **46**, 121–134.
- Lara E, Mitchell EAD, Moreira D, López-García P (2011) Highly Diverse and Seasonally Dynamic Protist Community in a Pristine Peat Bog. *Protist*, **162**, 14–32.
- Lara E, Moreira D, López-García P (2010) The Environmental Clade LKM11 and Rozella Form the Deepest Branching Clade of Fungi. *Protist*, **161**, 116–121.
- Lara E, Moreira D, Vereshchaka A, López-García P (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplomonads. *Environmental microbiology*, **11**, 47–55.
- Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, **23**, 2947–8.
- Leander BS, Kuvardina ON, Aleshin V V., Mylnikov AP, Keeling PJ (2003) Molecular phylogeny and surface morphology of *Colpodella edax* (Alveolata): insights into the phagotrophic ancestry of apicomplexans. *The Journal of eukaryotic microbiology*, **50**, 334–40.
- Lecointre G, Le Guyader H (2006) *Classification phylogénétique du vivant. 3ème édition*. Paris, France.
- Lee RE (2008) *Phycology*. Cambridge University Press, Cambridge, UK.
- Lefèvre E, Bardot C, Noël C *et al.* (2007) Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environmental microbiology*, **9**, 61–71.
- Lefèvre E, Roussel B, Amblard C, Sime-Ngando T (2008) The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PloS One*, **3**, 10.
- Lefranc M, Thénot A, Lepère C, Debroas D (2005) Genetic diversity of small eukaryotes in lakes differing by their trophic status. *Applied and environmental microbiology*, **71**, 5935–5942.
- Legrand C, Johansson N, Johnsen G, Borsheim KY, Granéli E (2001) Phagotrophy and toxicity variation in the mixotrophic *Prymnesium patelliferum* (Haptophyceae). *Limnology and oceanography*, **46**, 1208–1214.
- Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature reviews. Microbiology*, **9**, 119–30.
- Lepère C, Boucher D, Jardillier L, Domaizon I, Debroas D (2006) Succession and regulation factors of small eukaryote community composition in a lacustrine ecosystem (Lake Pavin). *Applied and environmental microbiology*, **72**, 2971–2981.
- Lepère C, Domaizon I, Debroas D (2008) Unexpected importance of potential parasites in the composition of the freshwater small-eukaryote community. *Applied and environmental microbiology*, **74**, 2940–9.
- Lepère C, Domaizon I, Taib N *et al.* (2013) Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems. *FEMS microbiology ecology*, **85**, 85–94.



- Lepère C, Masquelier S, Mangot J-F, Debroas D, Domaizon I (2010) Vertical structure of small eukaryotes in three lakes that differ by their trophic status: a quantitative approach. *The ISME journal*, **4**, 1509–19.
- Li WKW (1994) Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnology and Oceanography*, **39**, 169–175.
- Lindmark DG, Müller M (1973) Hydrogenosome, a Cytoplasmic Organelle of the Anaerobic Flagellate *Tritrichomonas foetus*, and Its Role in Pyruvate Metabolism. *The Journal of biological chemistry*, **248**, 7724–7728.
- Liu H, Aris-Brosou S, Probert I, de Vargas C (2010) A time line of the environmental genetics of the haptophytes. *Molecular biology and evolution*, **27**, 161–76.
- Liu H, Probert I, Uitz J *et al.* (2009) Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12803–12808.
- Logares R, Brate J, Bertilsson S *et al.* (2009) Infrequent marine-freshwater transitions in the microbial world. *Trends in Microbiology*, **17**, 414–422.
- Logares R, Shalchian-Tabrizi K, Boltovskoy A, Rengefors K (2007) Extensive dinoflagellate phylogenies indicate infrequent marine-freshwater transitions. *Molecular phylogenetics and evolution*, **45**, 887–903.
- López-García P, Moreira D (2008) Tracking microbial biodiversity through molecular and genomic ecology. *Research in microbiology*, **159**, 67–73.
- López-García P, Philippe H, Gail F, Moreira D (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 697–702.
- López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, **409**, 603–7.
- Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 11436–40.
- Luo W, Bock C, Li H, Padisák J, Krienitz L (2011) Molecular and microscopic diversity of planktonic eukaryotes in the oligotrophic Lake Stechlin (Germany). *Hydrobiologia*, **661**, 133–143.
- Mangot J-F, Debroas D, Domaizon I (2010) Perkinsozoa, a well-known marine protozoan flagellate parasite group, newly identified in lacustrine systems: a review. *Hydrobiologia*, **659**, 37–48.
- Mangot J-F, Domaizon I, Taib N *et al.* (2012) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environmental microbiology*, **15**, 1745–58.
- Mangot J-F, Lepère C, Bouvier C, Debroas D, Domaizon I (2009) Community structure and dynamics of small eukaryotes targeted by new oligonucleotide probes: new insight into the lacustrine microbial food web. *Applied and environmental microbiology*, **75**, 6373–81.
- Manton I, Peterfi FRS, Peterfi LS (1969) Observations on the fine structure of coccoliths, scales and the protoplast of a freshwater coccolithophorid, *Hymenomonas roseola* Stein, with



supplementary observations on the protoplast of *Cricosphaera carterae*. *Proceedings of the Royal Society*, **172**, 1–15.

Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–80.

Martiny JBH, Bohannan BJM, Brown JH *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nature reviews. Microbiology*, **4**, 102–12.

Massana R (2011) Eukaryotic picoplankton in surface oceans. *Annual review of microbiology*, **65**, 91–110.

Massana R, Balagué V, Guillou L, Pedrós-Alió C (2004a) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS microbiology ecology*, **50**, 231–43.

Massana R, Del Campo J, Sieracki ME, Audic S, Logares R (2013) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *The ISME journal*.

Massana R, Castresana J, Balagué V *et al.* (2004b) Phylogenetic and ecological analysis of novel marine stramenopiles. *Applied and environmental microbiology*, **70**, 3528–34.

Massana R, Guillou L, Diez B, Pedrós-Alió C (2002) Unveiling the Organisms behind Novel Eukaryotic Ribosomal DNA Sequences from the Ocean. *Applied and Environmental Microbiology*, **68**, 4554–4558.

Massana R, Pedrós-Alió C (2008) Unveiling new microbial eukaryotes in the surface ocean. *Current opinion in microbiology*, **11**, 213–218.

Massana R, Terrado R, Forn I, Lovejoy C, Pedrós-Alió C (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environmental microbiology*, **8**, 1515–22.

McQuoid MR, Godhe A, Nordberg K (2002) Viability of phytoplankton resting stages in the sediments of a coastal Swedish fjord. *European Journal of Phycology*, **37**, 191–201.

Metzker M (2009) Sequencing technologies—the next generation. *Nature Reviews Genetics*, **11**, 31–46.

Monchy S, Sanciu G, Jobard M *et al.* (2011) Exploring and quantifying fungal diversity in freshwater lake ecosystems using rDNA cloning/sequencing and SSU tag pyrosequencing. *Environmental microbiology*, **13**, 1433–53.

Moon-van der Staay SY, van der Staay GW, Guillou L *et al.* (2000) Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. *Limnology and Oceanography*, **45**, 98–109.

Moon-van der Staay SY, Tzeneva VA, van der Staay GWM *et al.* (2006) Eukaryotic diversity in historical soil samples. *FEMS microbiology ecology*, **57**, 420–8.

Moon-van der Staay SY, Wachter R De, Vaultot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, **409**, 607–610.

- Moreira D, von der Heyden S, Bass D *et al.* (2007) Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Molecular phylogenetics and evolution*, **44**, 255–66.
- Moreira D, López-García P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends in Microbiology*, **10**, 31–38.
- Moreira D, López-García P (2014) The rise and fall of Picobiliphytes: How assumed autotrophs turned out to be heterotrophs. *BioEssays : news and reviews in molecular, cellular and developmental biology*.
- Moritz H (2000) Geodetic Reference System 1980. *Journal of Geodesy*, **74**, 128–133.
- Nakai R, Abe T, Baba T *et al.* (2012) Eukaryotic phylotypes in aquatic moss pillars inhabiting a freshwater lake in East Antarctica, based on 18S rRNA gene analysis. *Polar Biology*, **35**, 1495–1504.
- Nicholls KH (2003) Haptophyte algae. In: *Freshwater Algae of North America: Ecology and Classification* (eds Wehr JG, Sheath RG), pp. 511–521. Elsevier, San Diego, CA, USA.
- Nicholls KH, Beaver JL, Estabrook RH (1982) Lakewide odours in Ontario and New Hampshire caused by *Chrysochromulina breviturrita* Nich. (Prymnesiophyceae). *Hydrobiologia*, **96**, 91–95.
- Nolte V, Pandey R, Jost S (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Molecular ecology*, **19**, 2908–2915.
- Not F, Gausling R (2007) Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environmental microbiology*, **9**, 1233–52.
- Not F, Valentin K, Romari K, Lovejoy C (2007) Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science*, **315**, 253–5.
- Oertli B, Frossard P-A (2013) *Mares et étangs: Ecologie, gestion, aménagement et valorisation*. Presses polytechniques et universitaires romandes.
- Oikonomou A, Katsiapi M, Karayanni H, Moustaka-Gouni M, Kormas KA (2012) Plankton microorganisms coinciding with two consecutive mass fish kills in a newly reconstructed lake. *TheScientificWorldJournal*, **2012**, 504135.
- Okamoto N, Chantangsi C, Horák A, Leander BS, Keeling PJ (2009) Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the Hacrobia taxon nov. *PloS one*, **4**, e7080.
- Oksanen J, Blanchet FG, Kindt R *et al.* (2013) vegan: Community Ecology Package.
- Orsi W, Song YC, Hallam S, Edgcomb V (2012) Effect of oxygen minimum zone formation on communities of marine protists. *The ISME journal*, **6**, 1586–601.
- Pawlowski J (2013) The new micro-kingdoms of eukaryotes. *BMC biology*, **11**, 40.
- Pedrés-Alió C (2006) Marine microbial diversity: can it be determined? *Trends in microbiology*, **14**, 257–63.
- Pernthaler J (2005) Predation on prokaryotes in the water column and its ecological implications. *Nature reviews. Microbiology*, **3**, 537–46.

- Philippe H, Germot A, Moreira D (2000) The new phylogeny of eukaryotes. *Current Opinion in Genetics & Development*, **10**, 596–601.
- Pielou EC (1966) The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, **13**, 131–144.
- Piwosz K, Wiktor JM, Niemi A, Tatarek A, Michel C (2013) Mesoscale distribution and functional diversity of picoeukaryotes in the first-year sea ice of the Canadian Arctic. *The ISME journal*, **7**, 1461–71.
- Preisig HR (2002) Phylum Haptophyta (Prymnesiophyta). In: *The freshwater algal flora of the British isles: An identification guide to freshwater and terrestrial algae* (eds John DM, Whitton BA, Brook AJ), pp. 211–213. The press syndicate of the University of Cambridge, Cambridge, UK.
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. (AFY Poon, Ed.). *PloS one*, **5**, e9490.
- Pruesse E, Quast C, Knittel K *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research*, **35**, 7188–96.
- Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**, 38.
- R Development Core Team (2013) R: A language and environment for statistical computing.
- Ragon M, Fontaine MC, Moreira D, López-García P (2012) Different biogeographic patterns of prokaryotes and microbial eukaryotes in epilithic biofilms. *Molecular ecology*, **21**, 3852–68.
- Rappé MS, Suzuki MT, Vergin KL, Giovannoni SJ (1998) Phylogenetic Diversity of Ultraplankton Plastid Small-Subunit rRNA Genes Recovered in Environmental Nucleic Acid Samples from the Pacific and Atlantic Coasts of the United States. *Appl. Envir. Microbiol.*, **64**, 294–303.
- Rees GN, Watson GO, Baldwin DS, Mitchell AM (2006) Variability in sediment microbial communities in a semipermanent stream: impact of drought. *Journal of the North American Benthological Society*, **25**, 370–378.
- Reinfelder JR (2011) Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Annual review of marine science*, **3**, 291–315.
- Rengefors K, Weyhenmeyer GA, Bloch I (2012) Temperature as a driver for the expansion of the microalga *Gonyostomum semen* in Swedish lakes. *Harmful Algae*, **18**, 65–73.
- Richards TA, Bass D (2005) Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Current opinion in microbiology*, **8**, 240–52.
- Richards TA, Vepritskiy AA, Gouliamova DE, Nierzwicki-Bauer SA (2005) The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. *Environmental microbiology*, **7**, 1413–25.
- Romari K, Vaultot D (2004) Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnology and Oceanography*, **49**, 784 – 798.

- Scamardella JM (1999) Not plants or animals: a brief history of the origin of Kingdoms Protozoa, Protista and Protoctista. *International Microbiology*, **2**, 207–216.
- Scheckenbach F, Hausmann K, Wylezich C, Weitere M, Arndt H (2010) Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 115–20.
- Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, **75**, 7537–41.
- Scholin CA, Gulland F, Doucette GJ *et al.* (2000) Mortality of sea lions along the central California coast linked to a toxic diatom bloom. *Nature*, **403**, 80–4.
- Seenivasan R, Sausen N, Medlin LK, Melkonian M (2013) Picomonas judraskeda gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as “picobiliphytes”. (RF Waller, Ed.). *PloS one*, **8**, e59565.
- Shade A, Peter H, Allison SD *et al.* (2012a) Fundamentals of microbial community resistance and resilience. *Frontiers in microbiology*, **3**, 417.
- Shade A, Read JS, Youngblut ND *et al.* (2012b) Lake microbial communities are resilient after a whole-ecosystem disturbance. *The ISME journal*, **6**, 2153–67.
- Shalchian-Tabrizi K, Reier-Røberg K, Ree DK, Klaveness D, Bråte J (2011) Marine–freshwater colonizations of haptophytes inferred from phylogeny of environmental 18S rDNA sequences. *Journal of Eukaryotic Microbiology*, **58**, 315–8.
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature biotechnology*, **26**, 1135–45.
- Shi X, Marie D, Jardillier L, Scanlan D, Vaultot D (2009) Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLoS One*, **4**, e7657.
- Simó R (2001) Production of atmospheric sulfur by oceanic plankton: biogeochemical, ecological and evolutionary links. *Trends in Ecology & Evolution*, **16**, 287–294.
- Simon M, Jardillier L, Deschamps P *et al.* Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems. *Environmental Microbiology*.
- Simon M, López-García P, Moreira D, Jardillier L (2013) New haptophyte lineages and multiple independent colonizations of freshwater ecosystems. *Environmental microbiology reports*, **5**, 322–32.
- Simpson EH (1949) Measurement of Diversity. *Nature*, **163**, 688–688.
- Simpson AGB, Roger A (2004) The real “kingdoms” of eukaryotes. *Current Biology*.
- Šlapeta J, López-García P, Moreira D (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Molecular biology and evolution*, **23**, 23–9.
- Šlapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. *Proceedings of the Royal Society B Biological Sciences*, **272**, 2073–2081.

- Smith VH, Foster BL, Grover JP *et al.* (2005) Phytoplankton species richness scales consistently from laboratory microcosms to the world's oceans. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4393–6.
- Sogin ML, Elwood HJ, Gunderson JH (1986) Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proceedings of the National Academy of Sciences*, **83**, 1383–1387.
- Sogin ML, Morrison HG, Huber JA *et al.* (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12115–20.
- Soltys BJ, Gupta RS (1994) Presence and Cellular Distribution of a 60-kDa Protein Related to Mitochondrial HSP60 in *Giardia lamblia*. *The Journal of Parasitology*, **80**, 580–590.
- Søndergaard M, Jeppesen E, Jensen JP (2005) Pond or lake: does it make any difference? *Archiv für Hydrobiologie*, **162**, 143–165.
- Sonntag B, Posch T, Klammer S, Teubner K, Psenner R (2006) Phagotrophic ciliates and flagellates in an oligotrophic, deep, alpine lake : contrasting variability with seasons and depths. *Aquatic microbial ecology*, **43**, 193–207.
- Stockner JG (1988) Phototrophic picoplankton: an overview from marine and freshwater ecosystems. *Limnology and Oceanography*, **33**, 765–775.
- Stoeck T, Behnke A, Christen R *et al.* (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC biology*, **7**, 72.
- Stoeck T, Taylor GT, Epstein S (2003) Novel Eukaryotes from the Permanently Anoxic Cariaco Basin (Caribbean Sea). *Applied and Environmental Microbiology*, **69**, 5656–5663.
- Taib N, Mangot J-F, Domaizon I, Bronner G, Debroas D (2013) Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights into the freshwater protist diversity. (SJ Green, Ed.). *PloS one*, **8**, e58950.
- Takishita K, Tsuchiya M, Kawato M *et al.* (2007) Genetic Diversity of Microbial Eukaryotes in Anoxic Sediment of the Saline Meromictic Lake Namako-ike (Japan): On the Detection of Anaerobic or Anoxic-tolerant Lineages of Eukaryotes. *Protist*, **158**, 51–64.
- Tovar J, Fischer A, Clark CG (1999) The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Molecular Microbiology*, **32**, 1013–1021.
- Triadó-Margarit X, Casamayor EO (2012) Genetic diversity of planktonic eukaryotes in high mountain lakes (Central Pyrenees, Spain). *Environmental microbiology*, **14**, 2445–56.
- Trigal C, Hallstan S, Johansson KSL, Johnson RK (2013) Factors affecting occurrence and bloom formation of the nuisance flagellate *Gonyostomum semen* in boreal lakes. *Harmful Algae*, **27**, 60–67.
- Unrein F, Gasol JM, Not F, Forn I, Massana R (2014) Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. *The ISME journal*, **8**, 164–76.
- De Vargas C, Aubry MP, Probert I, Young J (2007) Origin and evolution of coccolithophores: from coastal hunters to oceanic farmers. In: *Evolution of primary producers in the sea* (eds Falkowski PG, Knoll a H), pp. 251–258. Elsevier, London, UK.

- Vaulot D, Eikrem W (2008) The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS microbiology reviews*, **32**, 795–820.
- Vigil P, Countway P, Rose J *et al.* (2009) Rapid shifts in dominant taxa among microbial eukaryotes in estuarine ecosystems. *Aquatic Microbial Ecology*, **54**, 83–100.
- Walker G, Dorrell R, Schlacht A, Dacks J (2011a) The eukaryotes in 2011. *84.237.21.152*, 1–26.
- Walker G, Dorrell R, Schlacht A, Dacks J (2011b) Eukaryotic systematics: a user's guide for cell biologists and parasitologists. *Parasitology*, **138**, 1638–63.
- Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. *Nature reviews. Genetics*, **3**, 850–61.
- Whittaker RH (1969) New Concepts of Kingdoms of Organisms. *Science*, **163**, 150–160.
- De Wit R, Bouvier T (2006) “*Everything is everywhere, but, the environment selects*”; what did Baas Becking and Beijerinck really say? *Environmental microbiology*, **8**, 755–8.
- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, **74**, 5088–5090.
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, **87**, 4576–4579.
- Wright RT, Coffin RB (1984) Measuring microzooplankton grazing on planktonic marine bacteria by its impact on bacterial production. *Microbial ecology*, **10**, 137–49.
- Yuan CL, Keeling PJ, Krause PJ *et al.* (2012) *Colpodella spp.*-like parasite infection in woman, China. *Emerging infectious diseases*, **18**, 125–7.
- Zhao B, Chen M, Sun Y, Yang J, Chen F (2011) Genetic diversity of picoeukaryotes in eight lakes differing in trophic status. *Canadian journal of microbiology*, **57**, 115–26.
- Zinger L, Gobet A, Pommier T (2011) Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, **21**, 1878–96.
- Zubkov M V, Tarran GA (2008) High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature*, **455**, 224–6.
- Zuckerkandl E, Pauling L (1965) Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, **8**, 357–366.



Ecosystème	Mare Gabard	Saint Robert	Etang des Vallées		Ru Sainte Anne	La Claye
Type	Mare forêt	Mare village	Etang		Ruisseau	Mare dans complexe
Fraction de taille	0,2-5µm	0,2-5µm	0,2-5µm	5-30µm	0,2-5µm	0,2-5µm
avril-2011	MG3	SR3	EV11	EV12	RSA3	LC3
mai-2011	MG5	SR5	EV13	EV14	RSA5	LC5
juin-2011	MG7	SR7	EV15	EV16	RSA7	LC7
juillet-2011	MG9	SR9	EV17	EV18	RSA9	<b>LC9</b>
septembre-2011	MG11	SR11	EV19	EV20	RSA11	<b>LC11</b>
octobre-2011	MG13	SR13	EV21	EV22	RSA13	<b>LC13</b>
novembre-2011	MG15	SR15	EV23	EV24	RSA15	<b>LC15</b>
décembre-2011	MG17	SR17	EV25	EV26	RSA17	<b>LC17/LC17b</b>
janvier-2012	MG19	SR19	EV27	EV28	RSA19	LC19
février-2012	MG21	SR21	EV29	EV30	RSA21	LC21
mars-2012	MG23	SR23	EV31	EV32b/EV32	RSA23	LC23
avril-2012	MG25/MG25b	SR25/SR25b	EV33/EV33c	EV34/EV34b	RSA25/RSA25b	LC25/LC25b
mai-2012	MG27	SR27	EV35	EV36	RSA27	LC27
juin-2012	MG29	SR29	EV37	EV38	RSA29	LC29
juillet-2012	MG31	SR31	EV39	EV40	RSA31	LC31
août-2012	MG33	SR33	EV41	EV42	<b>RSA33</b>	LC33
septembre-2012	MG35	SR35	EV43	EV44	<b>RSA35</b>	<b>LC35</b>
octobre-2012	MG37	SR37	EV45	EV46	RSA37	LC37
novembre-2012	MG39	SR39	EV47	EV48	RSA39	LC39
décembre-2012	MG41	SR41	EV49	EV50	RSA41	LC41
janvier-2013	MG43	SR43	EV51	EV52	RSA43	LC43
février-2013	MG45	SR45	EV53	EV54	RSA45	LC45
mars-2013	MG47	SR47	EV55	EV56	RSA47	LC47
avril-2013	MG49	SR49	EV57	EV58	RSA49	LC49

**Annexe : Principales caractéristiques des échantillons prélevés au cours du suivi mensuel sur 2ans de la diversité générale des petits eucaryotes dans 4 mares et un ru (chapitres 2 à 4).** Chaque nom d'échantillon est formé de lettres indiquant l'écosystème et d'un nombre indiquant le mois où il a été prélevé. Les échantillons d'avril 2011 à juillet 2011 ont été prélevés en fin de mois, les autres l'ont été en tout début du mois indiqué.

Les couleurs des cases "date" correspondent à la saison:

Printemps	Eté
Automne	Hiver

Les échantillons de **sédiment** (écosystème à sec, pas d'échantillon de plancton à cette date donc) sont indiqués en gras et en marron. Ils ne sont intégrés aux analyses qu'au chapitre 4.

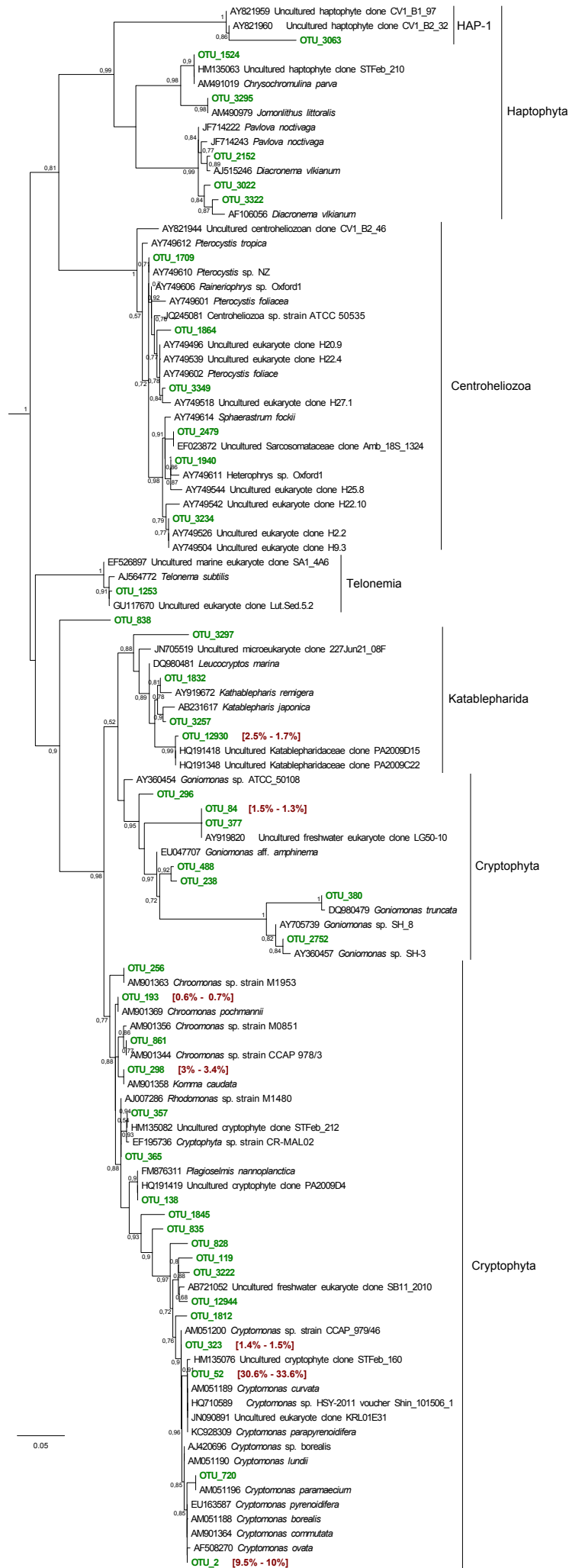
Deux noms d'échantillons dans une même case correspondent à des répliquats. Les échantillons analysés au chapitre 2 (avril 2012) sont tous réalisés en répliquats. Pour les analyses du suivi seul le répliquat ayant fourni le plus de séquences (indiqué en noir) est utilisé.



***Annexe 2 :***

*Reconstitutions phylogénétiques issues du matériel supplémentaires de l'article 2 (Chapitre 2).*

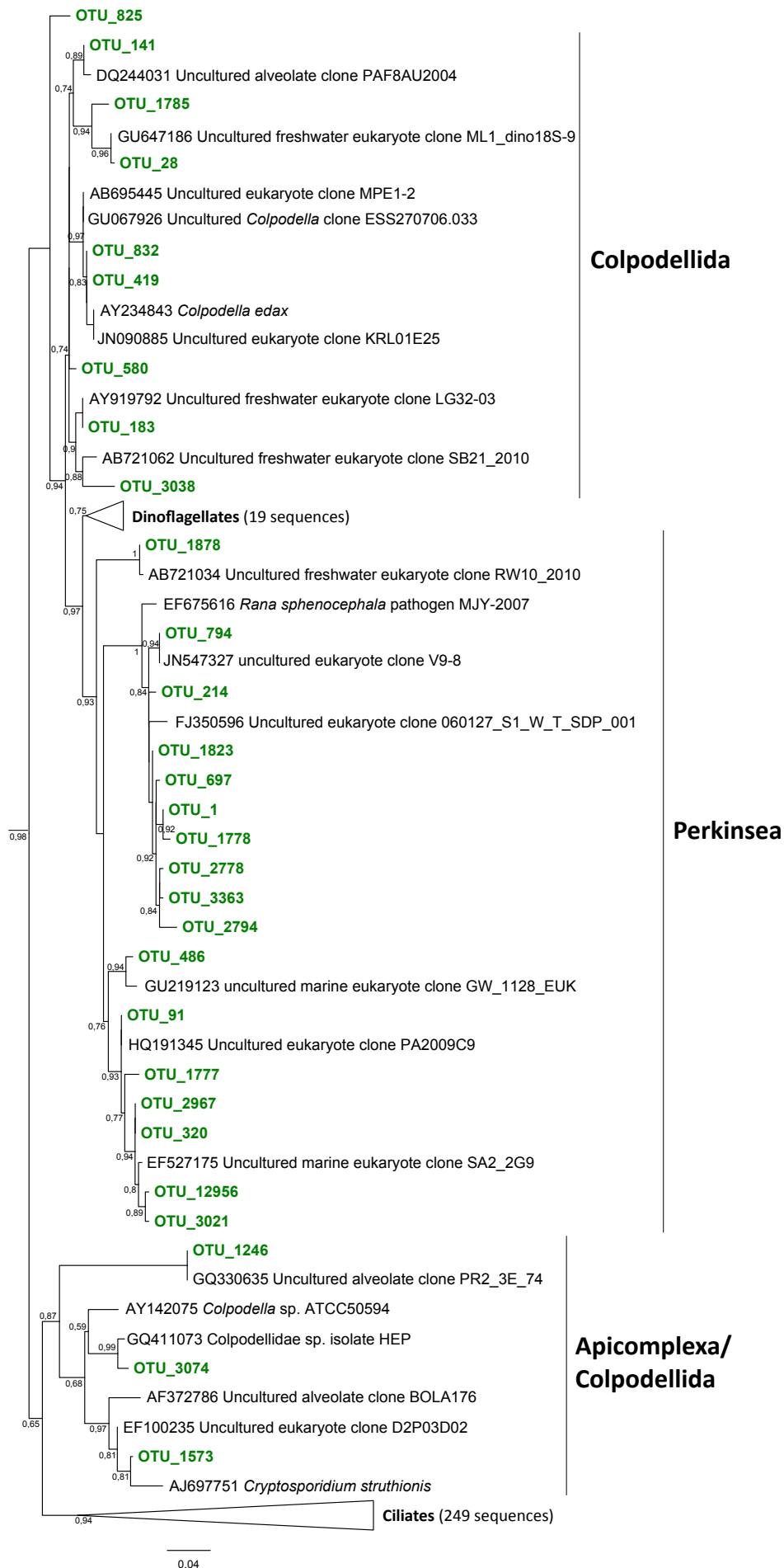
*Elles ne sont pas intégrées au document imprimé (non lisibles en A4)*



**Figure S2.** Approximate Maximum Likelihood (ML) phylogenetic tree of partial 18S rRNA gene sequences of cryptophytes, haptophytes, telonemids, katablepharids and centroheliozoans. A total of 354 unambiguously aligned positions were used to reconstruct the tree. 2 chlorophyte sequences were used as outgroup (not shown) to root the tree. Representative sequences of OTUs from this work are shown in bold green. For abundant OTUs (> 0.5% reads), their proportions in term of total number of reads and reads from the 0.2-5 µm size-fraction are given within square brackets. Statistical local support values higher than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.

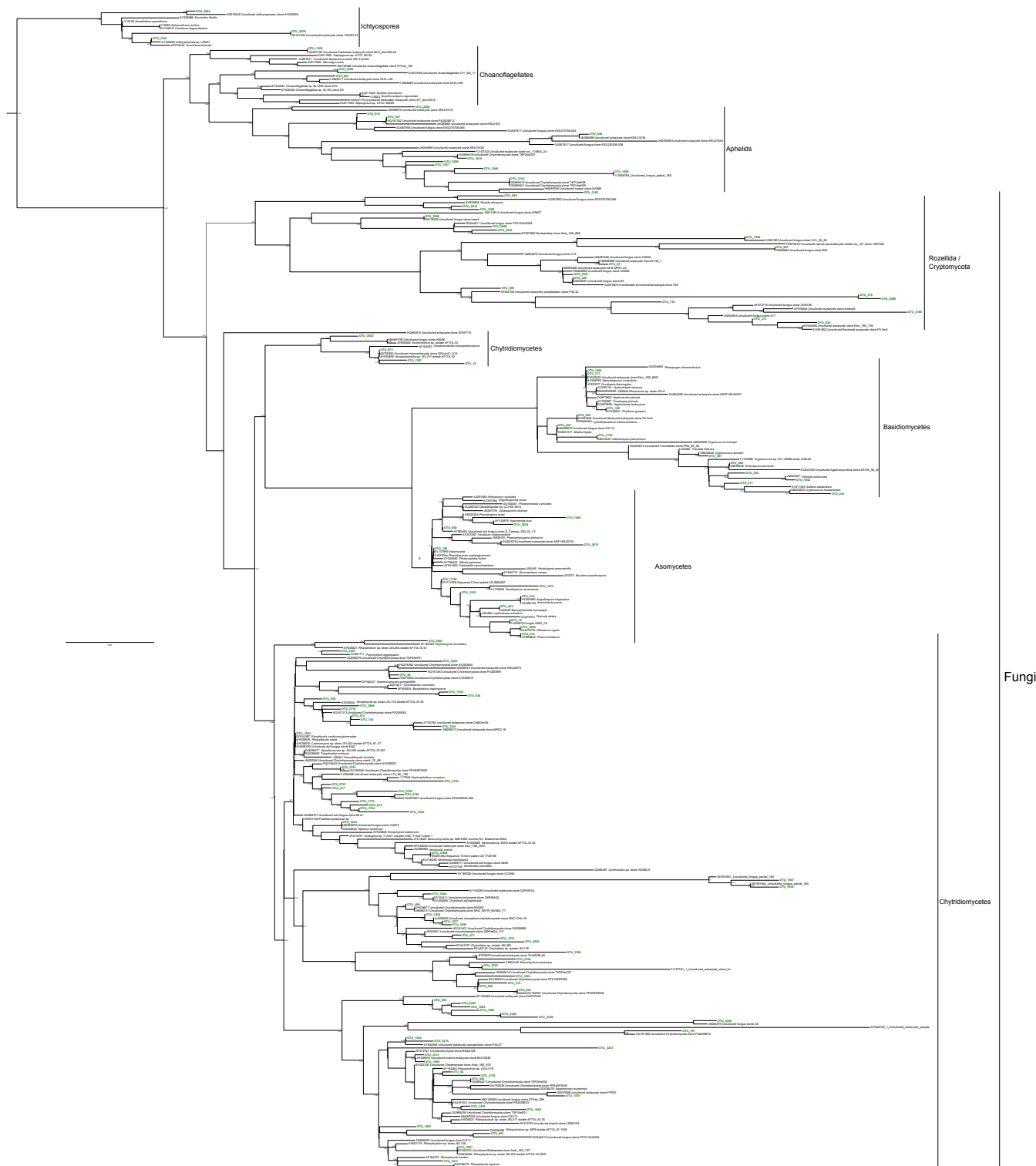


Figure S3. Phylogenetic tree of 102 eukaryotic species. A total of 102 eukaryotic species were used to reconstruct the tree. The bootstrap support values are shown at the nodes. Bootstrap values of 100% are shown at the nodes. Local support values greater than 0.5 are shown at nodes. The scale bar represents the number of substitutions per position for a unit branch length.

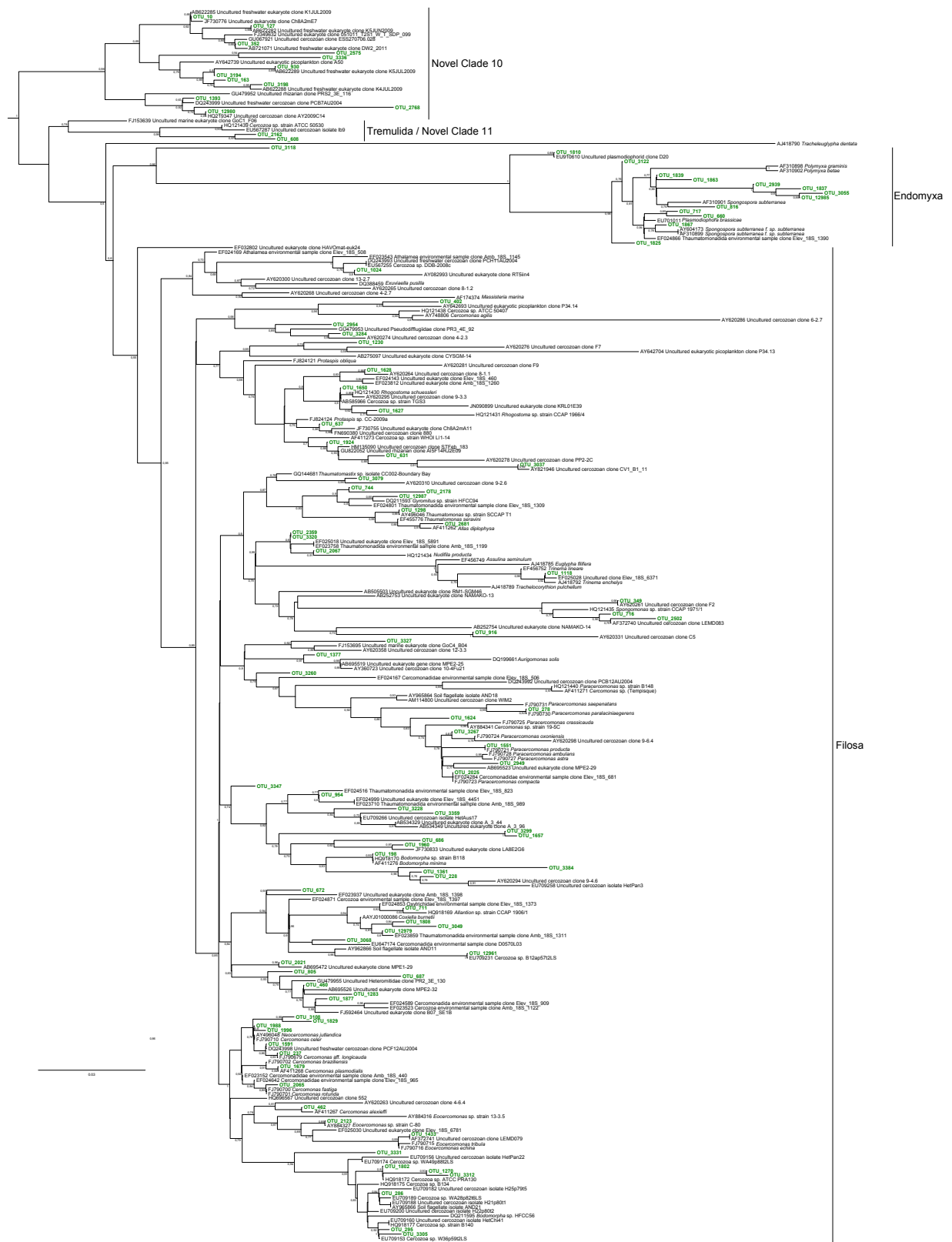


**Figure S4.** Approximate ML phylogenetic tree of partial 18S rRNA gene sequences of alveolates. Dinoflagellate and ciliate branches are shown collapsed. A total of 325 unambiguously aligned positions were used to reconstruct the tree. Four stramenopile sequences were used as outgroup to root the tree. Representative sequences of OTUs from this work are shown in bold green. Local support values higher than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.



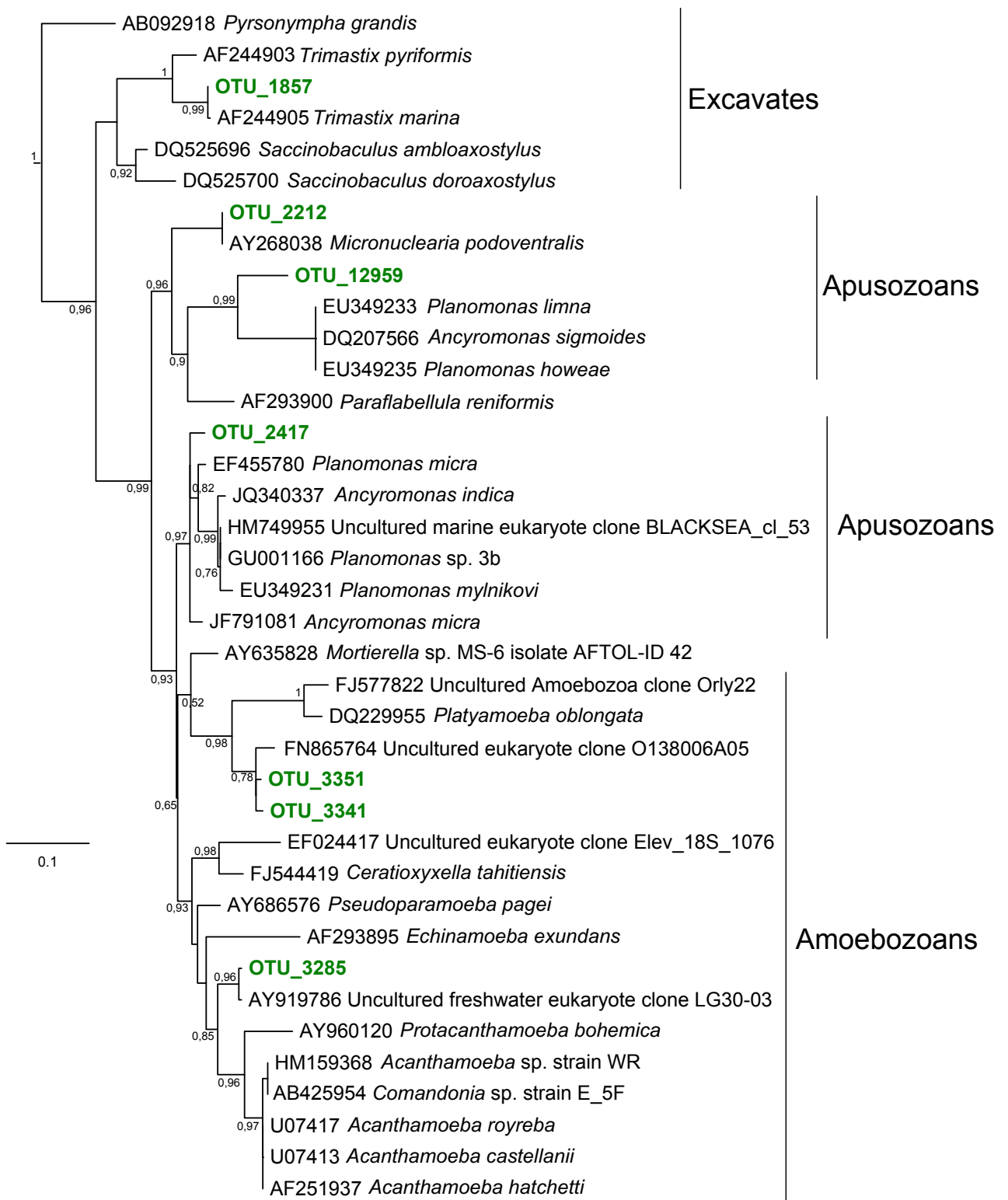


**Figure S6.** Approximate ML phylogenetic tree of 18S rDNA partial sequences of opisthokonts. A total of 342 unambiguously aligned positions were used to reconstruct the tree. Two amoebozoan sequences were used as outgroup to root the tree. Representative OTU sequences from our work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit of branch length.









**Figure S9.** Approximate ML phylogenetic tree of partial 18S rDNA sequences of excavates, apusozoans and amoebozoans. A total of 339 unambiguously aligned positions were used to reconstruct the tree. Two bacterian sequences were used as an outgroup to root the tree. Representative sequences of OTUs from that work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit branch length.

## Diversité, distribution spatiale et dynamique temporelle des petits eucaryotes dans des écosystèmes d'eau douce peu profonds

La diversité des très petits eucaryotes ( $<5\ \mu\text{m}$ ) a essentiellement été étudiée par des méthodes moléculaires dans les océans ou de grands lacs. La diversité dans les écosystèmes d'eau douce peu profonds et/ou étendus reste très peu explorée, bien que ces systèmes soient très nombreux et écologiquement importants dans les régions tempérées. Dans ce travail, nous avons voulu explorer la diversité et certains aspects de l'écologie des micro-organismes eucaryotes dans ce type d'écosystèmes, à l'aide de méthodes moléculaires ciblant l'ADNr 18S de cellules planctoniques de surface, dans la fraction de taille théorique 0,2-5  $\mu\text{m}$ .

Nous nous sommes d'abord concentrés sur les haptophytes, un groupe important en milieu marin mais beaucoup moins bien connu en eaux douces. Nous avons exploré leur diversité à l'aide d'amorces spécifiques pour amplifier les gènes des ARNr 18S du groupe suivi du clonage / séquençage Sanger de ces marqueurs, dans 17 écosystèmes continentaux et 2 colonnes d'eau marines pour comparer la diversité dans différents milieux, ainsi qu'à l'aide du pyroséquençage de ce même marqueur dans 4 mares et 1 ru au cours d'un suivi mensuel sur 2 ans. La diversité des haptophytes était moindre en eau douce qu'en milieu marin, mais nous avons pu y détecter un nouveau groupe, divergeant au sein des Isochrysidales, présentant une saisonnalité marquée. Les phylotypes d'eau douce étaient majoritairement distincts de ceux détectés en milieu marin, et ont confirmé l'existence de plusieurs transitions marin/eau douce dans l'histoire des haptophytes.

Dans un second temps, nous avons exploré par pyroséquençage 454 des ADNr 18S la diversité des micro-organismes eucaryotes dans 4 mares et 1 ru, échantillonnés au printemps, et différant par leur taille, leur forme et leur environnement proche. Nous avons pu détecter une grande diversité dans chaque système étudié, avec des séquences affiliées à tous les supergroupes reconnus (Archaeplastida, Stramenopiles, Alveolata, Rhizaria, Excavata, Amoebozoa et Opisthokonta), ainsi qu'à des taxa de position phylogénétique mal résolue tels que les Cryptophyta, Haptophyta, Centroheliozoa et Katablepharida. Notamment, certaines OTU étaient affiliées au groupe MAST-3 (MARine STRamenopiles) jusque-là considéré comme exclusivement marin, et potentiellement à d'autres groupes dans le même cas. Les communautés de petits eucaryotes étaient différentes dans chacun des écosystèmes ; ces différences ne corrélaient pas avec les distances géographiques entre sites (test de Mantel), et des analyses multivariées n'ont pas mis en évidence de relation claire entre la distribution d'aucun groupe et un quelconque paramètre environnemental.

Par la suite, nous avons réalisé un suivi de la diversité des micro-organismes microbiens sur 2 ans dans les mêmes 5 écosystèmes. Nous avons ainsi collecté des échantillons de plancton et mesuré différents paramètres physico-chimiques chaque mois, sauf pour 2 des écosystèmes lorsqu'ils étaient totalement à sec. La diversité détectée sur 2 ans était bien plus grande que celle identifiée lors de l'étude ponctuelle précédente. Cryptophytes, ciliés, chrysophytes et champignons *stricto sensu* étaient globalement les plus abondants. La composition et la structure des communautés différaient d'un écosystème à l'autre sur l'ensemble du suivi. Ces communautés étaient très dynamiques, et montraient une saisonnalité claire. Nous avons réalisé des analyses statistiques multivariées associant composition des communautés et données physico-chimiques. La corrélation la plus claire associait la distribution spatio-temporelle des champignons *sensu stricto* et les hautes valeurs de conductivité.

En dernier lieu, nous avons décrit la dynamique des communautés de petits eucaryotes dans l'une des mares et le ru lors d'épisodes de sécheresse. Nous avons collecté du sédiment dans le lit asséché des écosystèmes lors des sécheresses, et du plancton le reste du temps. Les communautés du sédiment présentaient une signature différente des assemblages planctoniques. Ces derniers montraient une résilience élevée, et retrouvaient une signature planctonique moins d'un mois après que les écosystèmes soient de nouveau en eau.

Ainsi, les petits écosystèmes d'eau douce peu profonds présentent un fonctionnement très particulier, avec des variations spatiales et temporelles fortes et rapides des conditions environnementales. Ils abritent une diversité de petits eucaryotes jusqu'ici insoupçonnée. Ces écosystèmes constituent de plus d'excellents modèles pour l'étude de l'écologie des micro-organismes eucaryotes.

# Diversity, spatial distribution and temporal dynamics of small eukaryotes in shallow freshwater ecosystems

The diversity of very small eukaryotes (<5 µm) has mainly been studied by molecular methods in marine systems or in large lakes. However, that of small shallow systems remains practically unexplored, despite the fact that these systems are extensive and ecologically important in temperate regions. We thus aimed at describing the diversity and community composition of small eukaryotes in shallow freshwater systems, using molecular methods targeting the 18S rRNA gene of planktonic cells in the 0.2-5 µm size range.

We first focused on haptophytes, an important group in marine environments but much less known in freshwaters. We explored their diversity using newly designed specific primers to amplify haptophyte 18S rRNA genes, followed by their subsequent cloning and Sanger sequencing in seventeen continental ecosystems and in two marine water columns to allow comparisons between different environments, as well as using 454-pyrosequencing in 4 ponds and one brook during a 2-years monthly survey. Even if freshwater haptophytes were less diverse than marine lineages, we revealed the presence of a divergent lineage belonging to the Isochrysidales never recorded so far, which presented a marked seasonality. Freshwater phylotypes were usually distinct from their marine counterparts, and confirmed the occurrence of multiple marine–freshwater transitions in haptophyte evolution.

In a second step, we explored the microbial eukaryote diversity in 5 distinct shallow ecosystems sampled at spring and that differ in size, shape and surrounding environment, by 454-pyrosequencing their 18S rDNA. Diversity was high in the studied systems, with sequences affiliated to the 7 recognized eukaryotic supergroups (Archaeplastida, Stramenopiles, Alveolata, Rhizaria, Excavata, Amoebozoa and Opisthokonta) as well as groups of unresolved phylogenetic position including, among others, Cryptophyta, Haptophyta, Centroheliozoa or Katablepharida. Especially, we detected OTUs affiliated to the previously thought exclusively marine lineage MAST-3 (MARine STramenopiles), and potentially to other MAST groups with no known representative from freshwaters. Small eukaryote community structures were different in the five ecosystems. Differences in community compositions did not correlate with geographical distances (Mantel test), and multivariate statistical analyses did not reveal clear relationships between any group distribution and specific environmental parameters.

Then, we conducted a 2-years survey of eukaryotic micro-organisms diversity in the same 5 small ecosystems. To do so, we collected plankton and measured several physical and chemical parameters on a monthly basis, except for two systems when they were totally dry. The total diversity encountered during the 24 months was much broader than that identified in the previous snapshot study. The most abundant detected groups were Cryptophytes, Ciliates, Chrysophytes and Fungi *sensu stricto*. Community structures and compositions were different in the five systems along the two years. In all systems, communities were highly dynamic, and revealed a marked seasonality, notably with summer and winter communities being always distinct. Multivariate statistical analyses were used to analyze simultaneously physico-chemical data and community compositions. The clearest correlation associated fungi distribution and high conductivity.

Finally, we described the dynamics of small-eukaryote communities in a pond and a brook through drought events. We collected sediment on the system beds when they were dry, and plankton the rest of time. Communities in the sediment and in the water presented distinct signatures. Surface water communities presented (a high) resilience, and recovered a planktonic signature within a month after the systems were filled up again with water.

In conclusion, small and shallow freshwater ecosystems show high and rapid spacio-temporal variations of environmental parameters. They host an unsuspected molecular diversity of small eukaryotes. Furthermore, they could constitute particularly good models to question important aspects of microbial ecology.